AGING, RELATIVE NUMEROUSNESS JUDGMENTS, AND SUMMATION IN WESTERN LOWLAND GORILLAS

A Thesis
Presented to
The Academic Faculty

by

Ursula S. Anderson

In Partial Fulfillment
Of the Requirement for the Degree
Master of Science in Psychology

Georgia Institute of Technology
November, 2003

AGING, RELATIVE NUMEROUSNESS JUDGMENTS,
AND SUMMATION IN WESTERN LOWLAND GORILLAS

Approved by:

Dr. Terry L. Maple, Advisor

Dr. M. Jackson Marr

Dr. Anderson D. Smith

Dr. Mollie A. Bloomsmith

Date Approved

ACKNOWLEDGMENTS

This thesis marks one milestone in my endeavor as a student of comparative psychology. Firstly, I would like to thank my family for their unending patience throughout a difficult writing process. In addition, I would like to thank my committee members, Dr's Terry Maple, Mollie Bloomsmith, Jack Marr, and Andy Smith for all their help, comments, concerns, and suggestions. Moreover Dr. Bloomsmith, for initially giving me the direction, "Aging", on which to focus.

I also thank Dr. Douglas Candland, my undergraduate advisor and mentor, who lit in me the fires of comparative psychology, ethology, and animal behavior, a three-headed beast that I thankfully cannot now extinguish. And the Rumbaugh's and other colleagues who did the initial empirical work that I followed.

Also, for allowing me to carry out this research I thank Zoo Atlanta and the School of Psychology at Georgia Institute of Technology. And to the curator of primates, Charles Horton, and the primate keeper staff, Bernie Gregory, Andrea Clay, Christine Mallar, and Laura Mayo, at Zoo Atlanta for facilitating data collection, thank you.

TABLE OF CONTENTS

ACKNOWLEGMENTS	iii
LIST OF TABLES	v
LIST OF ILLUSTRATIONS	vii
SUMMARY	ix
CHAPTER I INTRODUCTION	1
Learning, Memory, and Aging in Apes	
Numerical Competence in Apes	
Aging and Numerical Ability in Humans	
Objectives and Hypotheses	
CHAPTER II METHODS	25
Subjects	
Apparatus	
Procedure	
CHAPTER III RESULTS	36
Experiment 1	
Experiment 2	
Experiment 3	
CHAPTER IV DISCUSSION	93
Relative Numerousness Judgments Without Prior Specific Training	
Relative Numerousness Judgments With Specific Training	
Summation Without Prior Specific Training	
Relative Numerousness Judgments, Summation, and Subitizing	
Conclusions	
REFERENCES	109

LIST OF TABLES

Table 1	Subject name, sex, age, age group, birth year, birth type, and testing type	27
Table 2.1	Set of possible quantity comparisons and arrangements for experiment 1 and 2	33
Table 2.2	Set of possible quantity comparisons and arrangements for experiment 3	35
Table 3	Percent of correct responses and standard error (SE) each testing day	38
Table 4	Fisher's exact tests of frequency of correct responses to quantity variables by age group	40
Table 5	Binomial tests of the frequency of correct responses	46
Table 6	Chi-square tests of the frequency of correct responses each testing day	46
Table 7	Percent of correct responses and standard error (SE) each testing day	47
Table 8	Pearson correlations between percent of correct response and the quantity variables	50
Table 9	Percent of correct responses to the quantity comparisons	51
Table 10	Kruskal-Wallis mean rank and mean response time (RT)	53
Table 11	Wilcoxon signed-rank tests of response time (RT) between correct and incorrect responses	53
Table 12	Pearson correlations between response time, quantity variables, and percent of correct responses	54

Table 13	Fisher's exact tests of the frequency of correct responses to quantity variables by age group for the last 2 testing days	56
Table 14	Binomial tests of the frequency of correct responses	63
Table 15	Kruskal-Wallis ranks of the number of testing days and trials to criterion	64
Table 16	Percent of correct responses and standard error (SE) each testing day	65
Table 17	Binomial tests of the frequency of correct responses to quantity comparisons for all testing days	69
Table 18	Pearson correlations between percent of correct responses and the quantity variables for the last 2 testing days	71
Table 19	Kruskal-Wallis mean rank and mean response time (RT) for the last 2 testing days	72
Table 20	Pearson correlations between response time, quantity variables, and the percent of correct responses	73
Table 21	Wilcoxon signed-rank tests of response time (RT) between correct and incorrect responses	73
Table 22	Fisher's exact tests of the frequency of correct responses to the quantity variables by age group	75
Table 23	Binomial tests of the frequency of correct responses to quantity comparisons by age group	83
Table 24	Binomial tests of the frequency of correct responses	88
Table 25	Chi-square tests of the frequency of correct responses each	88

	testing day	
Table 26	Pearson correlations between percent of correct responses and the quantity variables	91
Table 27	Pearson correlations between response time and the quantity variables	92

LIST OF ILLUSTRATIONS

Figure 1	Testing Apparatus	27
Figure 2	Percent of correct responses each testing day	39
Figure 3	Percent of correct responses to quantity comparisons	42
Figure 4.1	Mean response time to ratios between quantities	43
Figure 4.2	Mean response time to differences between quantities	44
Figure 4.3	Mean response time to totals of quantities	45
Figure 5.1	Percent of correct responses for the young each testing day	48
Figure 5.2	Percent of correct responses for the old each testing day	49
Figure 6	Percent of correct responses to the quantity comparisons for the last 2 testing days	58
Figure 7.1	Mean response time to ratios between quantities for the last 2 testing days	60
Figure 7.2	Mean response time to differences between quantities for the last 2 testing days	61
Figure 7.3	Mean response time to totals of quantities for the last 2 testing days	62
Figure 8.1	Percent of correct responses for the young each testing day	66
Figure 8.2	Percent of correct responses for the old each testing day	67
Figure 9.1	Percent of correct responses to differences between quantities for the young	76
Figure 9.2	Percent of correct responses to differences between quantities for the old	77

Figure 10.1	Percent of correct responses to ratios between quantities for the young	78
Figure 10.2	Percent of correct responses to ratios between quantities for the old	79
Figure 11.1	Percent of correct responses to totals of quantities for young	80
Figure 11.2	Percent of correct responses to totals of quantities for the old	81
Figure 12.1	Response time to totals of quantities for the young and old	85
Figure 12.2	Response time to differences between quantities for the young and old	86
Figure 12.3	Response time to ratios between quantities for the young and old	87
Figure 13.1	Percent of correct responses for the young each testing day	89
Figure 13.2	Percent of correct responses for the old each testing day	90

SUMMARY

The effect of age on numerical competence is unexplored in apes. Numerical competence encompasses all numerical, number-related, and number-based abilities, capabilities, and skills of nonhuman animals. Previous literature supports that apes possess a variety of numerical competencies, like relative numerousness judgments and performing summation operations. Relative numerousness judgments are defined as “more or less” judgments about two or more quantities. Summation is operationally defined as consistently choosing one pair of quantities whose overall sum is greater than the sum of another pair. Summation involves performing operations on quantities and may be a precursor to more formal numerical abilities in humans. The process of summation is theorized to entail both relative numerousness judgment and subitizing (a perceptual process of determining numerousness and numerosity). This study reviewed the literature reported on aging, memory, and learning in apes and monkeys and the literature reported on relative numerousness judgments, subitizing, estimation, and summation in apes, monkeys, and humans. These empirical reports were used as the theoretical framework for exploring aging, relative numerousness, and summation in gorillas (*Gorilla gorilla gorilla*).

Three experiments were performed. Experiment 1 tested the ability of the gorillas to perform relative numerousness judgments without prior specific training to do so; experiment 2 examined the ability of gorillas to perform relative numerousness judgments with specific training to do so and at near-asymptotic levels of performance; experiment 3 tested the ability of the gorillas to perform summation operations without prior specific training to do so. The hypotheses

were that gorillas would be able to perform both relative numerousness judgments before and during training, and perform summation without training to do so. Also, I hypothesized that old gorillas would perform relative numerousness judgments and summation operations less reliably and more slowly than young gorillas.

The findings indicated that the gorillas did not perform relative numerousness judgments to choose the larger quantity until after specific training to do so. However, the gorillas did perform summation operations without having been specifically trained to do so. Also, old gorillas performed more poorly and more slowly than young gorillas during summation tasks. The pattern of age differences during summation tasks was such that in general old gorillas performed more poorly and more slowly than their younger counterparts, but for specific ratios, differences, and totals of/between quantities. The findings did not suggest evidence for subitizing during relative numerousness judgments or during summation. However, like the existing literature on relative numerousness and summation tasks, the ratio, difference, and total between/of quantities affected relative numerousness and summation performance. The implications from these experiments are that subject age should be considered as a variate in future investigations in cognition, specifically numerical competence in apes and that the performance of gorillas on relative numerousness and summation tasks are comparable to the performance of chimpanzees and orangutans.

CHAPTER I

INTRODUCTION

The effect of aging on nonhuman primate cognition is currently receiving considerable attention by researchers. However, the group of nonhuman primates the most phylogenetically proximate to humans, the great apes, receives relatively very little attention in cognitive aging. And when we focus our attention to the cognitive area of numerical competence, the relation between aging and numerical competence remains unexplored in both apes and monkeys. As such, the purpose of this scientific endeavor was to investigate age-related differences in numerical competence in a species of ape. And because age-related differences in numerical competence have not been empirically examined in nonhuman primates, the performance of older apes on learning and memory tasks and the performance of older humans on numerical and arithmetic tasks are discussed as the evaluative framework of aging and numerical competence in apes.

Learning, Memory, and Aging in Apes

There are only a few reports that detail age-related differences in the learning and memory functioning of apes. Monkeys are the preferred subjects and many age-related differences described in monkeys parallel those found in the human population for memory and learning (for review see, Walker et. al, 1988; Woodruff-Pak, 1990; Rapp, 1995; Albert & Ross, 1996; Gallagher & Rap, 1997). Of the few existing reports on the relation between aging and memory and learning in apes, none involve gorillas only chimpanzees (*Pan troglodytes*) are used as subjects. These few empirical accounts that examine learning and memory suggest that old chimpanzees do not exhibit global or universal impairments to learning and memory. Depending on the task, old chimpanzees may learn and remember better, equal, or poorer than their younger counterparts remember. Also, old chimpanzees exhibit impairments to working memory, while longer-term memory functioning remains relatively intact.

Bernstein (1961) examined learning and memory performance in young (11 to 19 years old) and old (28 to 40 years old) chimpanzees. Three tests were administered, but none revealed any age-related learning or memory differences. Bernstein (1961) concluded that old chimpanzees did not display more response rigidity than did their younger counterparts. In the first test designed to measure response variability, subjects were allowed to rotate a wheel to bring food items within reach. Depending on the location of the food item, either a clock-or counterclockwise turn would result in the food item traveling the shortest route to the subject. The performance of young and old chimpanzees did not differ on this task; most chimpanzees brought the food to them by the shortest route. The second test was an object discrimination task, and age-related differences were not found. Young and old chimpanzees responded similarly and without deteriorations in performance over three one-month retention periods, indicating that the old chimpanzees were just as able as the young to learn and retain new associations for long periods. Similarly, other authors have reported finding no difference in object discrimination performance between old and young monkeys (Rapp, 1990; Bachevalier et al., 1991; Lai, Moss, Killiany, Rosene, & Herndon, 1995) and retention without deterioration of the object discrimination set for 14 days in old monkeys (Bakner & Treichler, 1993).

The final learning test was composed of conditional (signaled) and unconditional (unsignaled) object discrimination and reversal tasks. Bernstein (1961) found that old and young chimpanzees performed similarly during both acquisition and reversal, which suggests that old chimpanzees are able to not only learn new associations, but also reverse those previously learned associations. The same effect was also described in monkeys for object discrimination reversal (Rapp, 1990; Lai et al., 1995; Herndon, Moss, Rosene, & Killiany, 1997). The results from Bernstein's (1961) three learning and memory tests suggest that aged chimpanzees do not possess impairments to learning and long-term memory ability.

Riopelle and Rogers (1965) studied chimpanzee subjects between the ages of 7 and 41 years and found significant age-related differences for two of six memory and learning tasks—a spatial delayed-response task and a four-choice oddity test. The delayed-response task assessed working memory by measuring the ability of subjects to remember stimuli over varying, but relatively short intervals of time. Imposition of a delay between stimulus and response separates these two events in time and requires the maintenance across time of the stimulus-response contingency. The spatial delayed-response task revealed significant performance impairments in the old chimpanzees at shorter delays, but old and young chimpanzees performed similarly at the long delay. That the old chimpanzees performed poorer at the shorter delays may indicate a type of working memory impairment. And because the longer delay depressed the performance of the young chimpanzees, while performance of the old chimpanzees remained relatively the same, Riopelle and Rogers (1965) concluded that the young chimpanzees exhibited an attentional rather than retentional impairment. Specifically, the young animals became more distracted during the longer intervals and thus their performance matched the old animals.

A significant learning impairment in the old subjects was also found in the four-choice oddity test. In the oddity test chimpanzees were required to displace the odd panel in an array. The authors found that performance on the oddity test declined with age. Lastly, old and young chimpanzees performed similarly in a familiar and novel item identification task, a consecutive and concurrent object discrimination task, and a pattern discrimination task. Old and young monkeys were found to perform similarly on a pattern discrimination task (Bartus, Dean, & Fleming, 1979; Rapp, 1990). Riopelle and Rogers (1965) concluded that the effect of age on learning and memory performance varied widely from task to task, but overall old chimpanzees were not impaired with regard to learning new associations.

Approximately twenty years after Bernstein (1961) and Riopelle and Rogers (1965), Bloomstrand (1986) tested the learning and memory performance of two old chimpanzees and two young chimpanzees. The two old chimpanzees used by Bloomstrand (1986) were the same subjects Bernstein (1961) and Riopelle and Rogers (1965) used 20 years earlier as young subjects. Five learning and memory tests were given—a two-choice consecutive object discrimination task, a four-choice oddity test, a two-choice object discrimination reversal task, a single stimulus-pair discrimination reversal task, and a spatial delayed-response task. The two-choice consecutive object discrimination task and the four-choice oddity test were replications from Riopelle & Rogers' (1965) original work, and the single stimulus-pair discrimination reversal task was a replication from Bernstein (1961).

These three replications allowed for longitudinal analyses of age-related differences.

A longitudinal trend towards improved performance with advancing age was found by Bloomstrand (1986) for the two-choice object discrimination task—the old chimpanzees performed better than their performance twenty years earlier in the original task by Riopelle and Rogers (1965). But cross-sectional age differences were not found between old and young chimpanzees on this discrimination task. One interpretation of these results is that the long-term retention of the stimulus-response associations for the old chimpanzees resulted in their improved performance twenty years later. However, the retention of these stimulus-response associations did not improve their performance beyond that of the young chimpanzees. A longitudinal and cross-sectional trend towards performance improvement with advancing age was also found for the oddity test. The old chimpanzees performed significantly better twenty years later during testing by Bloomstrand (1986) than their original performance during testing by Riopelle and Rogers (1965) and the old chimpanzees also performed significantly better than did their young chimpanzee counterparts. These results could be interpreted as the old subjects retaining the oddity set for over twenty years, and consequently performing better than they did twenty years ago and better than other young chimpanzees. Johnson and Davis (1973) and Davis (1985) found that old monkeys performed better than young monkeys on oddity set learning and old monkeys without practice retained an oddity learning set for two seven-year periods.

The third test, the two-choice object discrimination reversal task indicated that old subjects performed better than the young subjects did during acquisition, but performance did not differ after reversal between old and young chimpanzees. The results of the single-stimulus pair object discrimination reversal task indicate that the old subjects performed more poorly than the young subjects did during acquisition, but performance did not differ after reversal between old and young chimpanzees (Bloomstrand, 1986). The fifth test was a spatial delayed-response task and the results indicated that performance at the shorter delays was not different between the two age groups. As the delay interval increased, both old and young showed declines in performance, but the old chimpanzees performed worse than the young chimpanzees. These results indicated that the old chimpanzees possessed normal memory ability when required to retain information for very short intervals, but for longer periods they exhibited memory deficits. The deficit may represent a true memory dysfunction resembling short-term memory loss in aged humans (Medin, 1969; Flicker, Bartus, Cook, & Ferris, 1984).

The results from Bloomstrand's (1986) delayed-response task are exactly opposite of Riopelle and Rogers (1965) delayed-response task findings, however, these findings are comparable to those found in monkeys for delayed-response tests (Medin, 1969; Borkhuis, Davis, & Medin, 1971; Bartus, Fleming, & Johnson, 1978; Davis, 1978; Bartus, Dean, & Beer, 1980; Marriott & Abelson, 1980; Davis, Bennet, & Weisenburger, 1982; Walker et al., 1988; Bachevalier et al., 1991; Bachevalier, 1993). In general, Bloomstrand's (1986) five tasks indicated that old chimpanzees exhibit cognitive flexibility by being able to reverse previously learned associations. Also, impairment to working memory may exist in old chimpanzees, but the ability to retain information for longer periods was relatively unimpaired.

More recently, Beran, Pate, Richardson, and Rumbaugh (2000) longitudinally examined one 27 year-old chimpanzee's recognition of lexigrams after a retention interval of more than 20 years between initial training and testing. "Lana" responded correctly significantly more times than chance would allow for five of the seven test stimuli learned more than 20 years prior. The subject's recognition of lexigrams learned more than 20 years prior supports the theory that overall long-term memory functioning does not deteriorate with advancing age in chimpanzees.

Numerical Competence in Apes

Ecologically, nonhuman primates encounter survival problems in which solutions may depend on quantitative competence. For example, when foraging animals may need to determine the rate of food return from various locations (Gallistel, 1989) and animals also must be able to keep track of groups and objects over time and space (Wynn, 1998). Numerical competence involves the numerical, number-related, or number based abilities, capabilities, or skills of nonhuman species. Investigations of numerical competence in apes have provided evidence that apes possess a variety of number-related capabilities; the matching of presented quantities to numerals (Ferster, 1964; Hayes & Nissen, 1971; Matsuzawa, 1985; Boysen, 1993), numerical ordering (Boysen, Berntson, Shreyer, & Quigley, 1993; Biro & Matsuzawa, 1999), proportionality (Woodruff & Premack, 1981), counting (Boysen & Berntson, 1989; Boysen, 1993; Rumbaugh & Washburn, 1993; Boysen, 1997; Beran, Rumbaugh, & Savage-Rumbaugh, 1998), and relative numerousness judgments, subitizing, estimating, and performing operations on numbers (for a review of numerical competence in apes see, Boysen & Hallberg, 2000). However, the possible age-related differences on these numerical abilities have escaped experimental evaluation in both monkeys and apes.

Davis and Pérusse (1988) described four major processes subsumed under numerical competence—counting, subitizing, estimation and relative numerousness judgments[1]. And while not defined as a major process of numerical competence, Davis and Pérusse (1988) discussed the possibility of animals performing operations on numbers analogous to human arithmetic. Counting was regarded as a more complex process than relative numerousness judgments, subitizing, or estimation. The authors defined counting as a formal enumerative process used to discriminate the absolute number of a set of items (Davis & Pérusse, 1988). Most relevant to the focus of this discussion are empirical accounts concerning subitizing, estimation, relative numerousness judgments, and performing operations on numbers.

Davis and Pérusse (1988) describe subitizing as a form of pattern recognition and labeling used to rapidly assign numerical labels to small quantities of items. Subitizing is a descriptive term that reflects the direct apprehension of a number (Thomas & Lorden, 1993) and involves an accurate and rapid judgment of numerosities up to some certain maximum numerosity. The findings of Murofushi (1997) and Tomonaga and Matsuzawa (2002) are similar to the earliest reports in humans indicating subitizing when the exposure duration of numerosity arrays were limited. That is, shallow or zero response time and error rate slopes when arrays are defined by numerosities from one to four (Kaufman, Lord, Reese, & Volkmann, 1949; Mandler & Shebo, 1982; Miller, 1993). Subitizing, a more

---

[1] Thomas and Lorden (1993) proposed an alternate account of the major processes involved in numerical competence.

autonomic response for determining numerosities less than four was proposed to account for these discontinuities.

Murofushi (1997) examined one nine-year old chimpanzee's ability to correctly assign the Arabic numeral that corresponded to the number of items displayed in an array in a matching-to-sample paradigm with unlimited exposure to the sample and comparison stimuli. Examining response times for correct responses as a function of the number of stimulus items indicated that response time for correct responses were stable from numerosities one to three in sets of arrays containing up to seven numerosities. The same chimpanzee was subsequently tested by Tomonaga and Matsuzawa (2002), as well as human subjects, on a similar task that varied the exposure duration of sample and comparison stimuli. The pattern of results obtained for the chimpanzee during unlimited exposure of the sample and comparison stimuli were similar to those obtained twenty years earlier. That is, response times for correct responses were fast and relatively constant and the number of errors was relatively constant for numerosities from one to five. The results obtained for numerosities from one to five in the human subjects during unlimited exposure were identical to the chimpanzees. The results obtained for the chimpanzee during limited exposure to the sample and comparison stimuli revealed response time for correct responses was fast and relatively constant and the number of errors was relatively constant up to numerosities of four. In humans during limited exposure, response time for correct responses was fast and relatively constant for numerosities up to four, and accurate and relatively constant for numerosities up to five.

Estimation referred to the ability of assigning a numerical label to an array consisting of a large number of items without counting or enumeration and it may involve a complex post-counting process (Davis & Pérusse, 1988). Estimation occurs in humans when the array contains too many discrete items to recognize patterns, and/or the array is presented for too short a period to permit counting (von Glaserfeld, 1993). Estimation is characterized by slower response speeds and greater errors (Miller, 1993; Thomas & Lorden, 1988), as it involves approximate determinations of quantitative values. Kaufman and colleagues (1949) and Mandler and Shebo (1982) in accordance with their findings in humans during limited exposures proposed estimation to occur for numerosities greater than six. The pattern was such response time and error slopes increased for numerosities from four to six and then after numerosities of six, response time remained increased but constant as the number of elements increased in the array.

Murofushi (1997) and Tomonaga and Matsuzawa (2002) concluded that their chimpanzee subject during unlimited exposure to the sample and comparison stimuli was estimating numerosities from four or five to the second largest numerosity in the array and the largest numerosity was determined by magnitude estimation. The authors found that response time increased for numerosities from four or five to the second largest numerosity in the set and then decreased for the largest numerosity in the set. Tomonaga and Matsuzawa (2002) also concluded that their chimpanzee subject during limited exposure responded for numerosities greater than five similarly to the data described previously in humans during estimation. The author found that during limited exposure, response time and error rates increased from numerosities of four to six, and from six to nine, response time remained constant. The human subjects tested by Tomonaga and Matsuzawa (2002) during unlimited and limited exposure did not exhibit patterns consistent with estimation; they exhibited response times with straight increasing slopes from numerosities of five to the maximum numerosity in the array. This finding by Tomonaga and Matsuzawa (2002) is identical to the pattern reported by others in humans during unlimited exposure to numerosity arrays and may demonstrate counting (Kaufman et al., 1949; Mandler & Shebo, 1982)

Davis and Pérusse (1988) described relative numerousness judgments as a dichotomous judgment of numerical inequality that may be ordered in magnitude (e.g., a more versus less comparison). Relative numerousness judgments involve discriminating between or ordering of usually two, but sometimes more than two quantities. It is assumed the simplest number-related skill since these judgments do not require comprehension of absolute number. The term numerousness is applied to this term to denote that the cardinality principle (the final item tag or label used to represent the numerosity of a set) of items in a set are not determined by counting. Numerosity refers to determining the cardinality attribute of sets of items by counting (Stevens, 1951). Davis (1983) described relative numerousness judgments as the cornerstone of foraging behavior and fundamental to the life of animals. And Cooper (1984) proposed the possibility of a developmental link in infant humans between relative numerousness judgments, a rudimentary skill and later emerging more formal enumerative processes requiring absolute number.

The ability of chimpanzees to discriminate relative quantities of food, performing relative numerousness judgments, was demonstrated by Menzel (1960). The author presented two adult chimpanzees with four banana pieces varying in size and orientation. The chimpanzees chose the largest piece first to eat followed by the second largest, third largest, and then the last piece. Analysis of the subjects' choice behavior revealed a high positive correlation between the actual food size and choice, such that larger sized food was chosen before smaller sized.

The findings by Menzel (1960) suggest that chimpanzees are sensitive to the relative size of food items and will generally select the largest quantity. The procedures developed to test relative numerousness judgments and summation presented in this discussion rest on these results, that is when given the choice the larger quantity of food is more preferable than a smaller quantity.

Dooley and Gill (1977) demonstrated more versus less relative numerousness judgments with Lana, a five year old chimpanzee with three years of previous language training, but no formal number-related training. Lana was simultaneously presented with two quantities of food items, ranging from one to five items, as in a two-choice discrimination paradigm. The quantities of food served as both reward and the stimuli to be discriminated and the subject was allowed to select and then receive one of the two quantities. The findings indicated that Lana reliably selected the larger quantity more often than the smaller quantity and her accuracy was not diminished even when comparing quantities differing by a single item. Dooley and Gill (1977) noted that determining the larger quantity may have been based on the number, surface area, and/or mass of the food items.

In subsequent relative numerousness tests, Lana was presented with two sets of objects rather than quantities of food in a two-choice discrimination paradigm. Each of the two sets of objects could consist of one to five objects that varied in size to prevent labeling based on relative area. Lana was required to label with lexigrams either the set representing “more” or “less” depending on which was requested by the experimenter’s lexigram question. Lana was rewarded only for correct trials and training continued until she reached the performance criterion of 70% correct for more and less labeling within a testing session. Testing consisted of the experimenter randomly requesting the subject to label the two object sets as “more” or “less” and the subject was rewarded only for correct trials. During testing, Lana was able to label the larger or the smaller set of quantities. Most errors were made on larger ratios between quantities than smaller. Specifically, larger and smaller sets were labeled more accurately for quantity comparisons of 1:5 (ratio of 0.20), 1:4 (ratio of 0.25), 1:3 (ratio of 0.33), 2:5 (ratio of 0.40) and 1:2 and 2:4 (ratios of 0.50) versus quantity comparisons of 2:3 and 3:5 (ratios of 0.67), 3:4 (ratio of 0.75), and 4:5 (ratio of 0.80). The relative numerousness judgment process was also found to generalize to novel and more complex comparisons when comparisons were increased to also include six to ten items as Lana reliably labeled either the larger or smaller set of objects in the novel comparisons (Dooley & Gill, 1977).

Rumbaugh, Savage-Rumbaugh, and Hegel (1987) investigated relative numerousness judgments in two adult chimpanzees aged approximately fifteen years old. In a two-choice discrimination paradigm, subjects were simultaneously presented with two quantities of food and allowed to select and then receive one of the two quantities. The two quantities each ranged from zero to four and the quantity of food served as both the reward and the stimuli to be discriminated. The authors found that the chimpanzees reliably chose the largest quantity in the set of two quantities, thus making a relative numerousness judgment.

The consistent ability to select the larger quantity from smaller quantities, thus performing relative numerousness judgment, was also demonstrated in two chimpanzees aged 9 and 32 (Boysen & Berntson, 1995) and five chimpanzees aged 6, 12, 14, 14.5, and 35 (Boysen, Berntson, Hannan, & Cacioppo, 1996). The chimpanzees consistently chose the larger quantity of food even though the contingency was reversed such that choosing the larger quantity resulted in receiving the smaller. Relative numerousness judgments were also performed by these chimpanzees when Arabic numerals were substituted in place of the two quantities with the contingency still reversed. Thus, the chimpanzees received the quantity of food corresponding to the unselected Arabic numeral. However, when Arabic numerals were used as stimuli the chimpanzees learned the reversed contingency, that is, the chimpanzees reliably selected the smaller Arabic numeral to receive the larger quantity of food. Relative numerousness judgments were also demonstrated in two orangutans in a nearly identical experiment requiring the comparison of two quantities of food ranging from one to six. Two orangutans aged 22 and 19 reliably chose the smaller discrete quantity of food in order to receive the larger quantity from the experimenter, demonstrating not only relative numerousness judgments but, also their ability to learn the reversed contingency (Shumaker, Palkovich, Beck, Guagnano, & Morowitz, 2001).

Monkeys are also able to perform relative numerousness judgments of quantities and of numerosities. Under reversed contingencies in which subjects received the unselected quantity of food, three monkeys presented with the quantity comparison 1:4 (Silberberg & Fujita, 1996) and seven of eight monkeys presented with quantity comparisons of 1:2, 1:4, and 1:6 (Anderson, Awazu, & Fujita, 2000) reliably selected the smaller discrete quantity in order to receive the larger quantity from the experimenter after correction methods of non-reward for incorrect responses, imposing inter-trial intervals after incorrect responses, and/or subsequently representing the two quantities until the subject responded correctly.

In the absence of reversed-contingencies, monkeys are also able to perform relative numerousness judgments. Thomas, Fowlkes, and Vickery (1980) trained two monkeys by reinforcing only correct responses to choose the smaller of two numerosities that were illustrated as dots on two separate cards. Both monkeys were able to discriminate and choose the smaller quantity of elements for numerosity comparisons of 2:3, 3:4, 4:5, 5:6, 6:7, and 7:8. And only one monkey reliably chose the smaller on comparisons of 8:9. Terrell and Thomas (1990) demonstrated numerousness judgments in four monkeys. The monkeys were trained by reinforcing only selections of the polygon with the smaller number of sides. Brightness, surface area, and patterning cues were controlled. During training all four monkeys were able to reliably discriminate to the criterion the smaller number of sides for polygon comparisons defined by 3:7, 4:7, and 5:7 sides, and three monkeys reliably discriminated 6:7, two monkeys 7:8, and one monkey 8:9 polygon sides.

Relative numerousness judgments to discriminate between values of Arabic numerals have also been demonstrated in monkeys. Washburn and Rumbaugh (1991) allowed two monkeys to select one Arabic numeral from the two presented. Subjects were rewarded with the number of food items represented by the Arabic numeral they choose. Analyses indicated that both monkeys reliably selected the larger of the two Arabic numerals in sets consisting of zero to five numerals. And when the Arabic numeral set consisted of zero to nine, one monkey reliably selected the numeral with the larger value. In a nearly identical numerousness task with two monkey subjects, Olthof, Iden, and Roberts (1997) found both monkeys reliably selected the larger of the two Arabic numerals in sets consisting of comparisons between numerals of zero, one, three, five, seven, and nine.

Summation is considered by Davis and Pérusse (1988) as possibly performing operations on quantities and Rumbaugh and colleagues (1987) suggested that it may be a precursor to formal arithmetic operations humans. Summation requires subjects to reliably select a pair of quantities whose overall total is greater than the total of another pair of quantities for all possible pairs within a numerical range. To determine which one of the two pairs of quantities are the largest, pairs must be combined, pooled, or summed to represent the total pair quantity.

In summation tests, Rumbaugh et al. (1987) presented subjects with two pairs of quantities. Each quantity ranged from zero to four, with each pair on a separate tray, and each quantity in a separate well of its tray. Subjects were allowed to select a tray and then they received the pair of quantities the selected tray contained. The authors not only found that the chimpanzees consistently selected the pair of quantities with the greater combined total, but that they selected the larger total paired quantity more frequently at smaller ratios between the quantities than at larger ratios. Specifically the larger pair total was chosen more frequently for quantity comparisons of 2:3 (ratio of 0.67) and 3:4 (ratio of 0.75), than at ratios of 4:5 (ratio of 0.80) and 5:6 (ratio of 0.83). The ratio between the quantities describes the relation between numerical distance between quantities and the numerical size of quantities. The summation processes was also found to generalize to novel and more complex comparisons when they introduced the quantity five into the existing comparisons which added thirty-two novel quantity comparisons. The results suggested that the chimpanzees used summation to choose the larger pair total, and summation generalized to more novel and complex comparisons. The authors theorized that the chimpanzees were first, subitizing individual quantities in each well, and then through elementary relational summation combining the subitized quantities for each pair, and finally to select the largest total pair quantity comparing the total combined value for pairs using relative numerousness judgments.

Subsequent experiments by Rumbaugh, Savage-Rumbaugh, and Pate (1988), using the same chimpanzees as Rumbaugh et al. (1987), indicated that the ability to reliably perform summation operations was not based on the avoidance of the pair of quantities containing the smallest single amount or selection of the pair of quantities containing the largest single amount. Errors were primarily made by chimpanzees when the total pair quantities for the two trays differed by a value of one, which indicated that they were very similar.

Nor was selection of the largest total pair based on the avoidance of pairs of quantities containing zero or one values (Pérusse & Rumbaugh, 1990). Pérusse and Rumbaugh (1990) tested two adult chimpanzees aged eighteen years, the same subjects used by Rumbaugh et al. (1987), to investigate the role of zero and one as values. The authors presented subjects with two quantities (in two-wells) for comparison to test relative numerousness judgments, and also two pairs of quantities (in four-wells) for comparison to test summation ability. The findings indicated that the chimpanzees selected the larger quantity significantly more often for two-well comparisons (relative numerousness judgment) than for four-well (summation) when the ratio between quantities was large. Large ratios indicate smaller numerical distances and/or smaller quantity sizes. Also, both chimpanzees were able to choose the larger total pair even when a pair of quantities did not contain a one or zero value. The conclusion drawn was that the chimpanzees were using a combinatorial process, not avoiding pairs of quantities containing one or zero, to determine and select units of larger quantity during the four-well tests. Rumbaugh and colleagues (1987) further suggested that the summation operations performed may be a precursor to formal addition of numbers in humans. In support of this suggestion, the two human subjects the authors tested performed similarly to the chimpanzees when they were given the same test with the constraint that responses be made within two seconds to eliminate formal counting (Pérusse & Rumbaugh, 1990).

Beran (2001) found similar results in two adult chimpanzees presented with sequential relative numerousness and sequential summation tests. These chimpanzees were aged 26 and 29 and one was a subject of Rumbaugh et al. (1987), Rumbaugh et al. (1988), and Pérusse and Rumbaugh (1990). In the relative numerousness experiments, chimpanzees were sequentially presented with two quantities, such that the two quantities needing comparison could not be viewed simultaneously. The author found that the chimpanzees selected the larger quantity significantly more often than chance even though the quantities to be compared could not be viewed together. Also, in performing relative numerousness judgments both subjects selected the larger quantity more reliably at smaller ratios and smaller total quantities. And in summation experiments, chimpanzees were sequentially presented with two pairs of quantities, such that the pair of quantities that needed to be combined could not be viewed simultaneously. Unlike the simultaneous presentations used by Rumbaugh et al. (1987), Rumbaugh et al. (1988), and Pérusse and Rumbaugh (1990), sequentially presenting the pairs of quantities eliminated the possibility of perceptual features aiding the chimpanzees in determining the larger pair. For the tests of summation, the authors found that the chimpanzees selected the larger total paired quantities significantly more often than chance would allow even though the total paired quantities to be compared and the pair of quantities to be combined could not be viewed together. And again, like previous summation experiments, subjects selected the larger total paired quantity more frequently at smaller ratios and larger differences between the paired quantities than at larger ratios and smaller differences. These results indicate that the subjects were mentally comparing and combining quantities through a type of representation since perceptual information was not available.

Further evidence that apes possess the ability to perform summation operations was advanced in the evaluation of estimation and comparison of discrete quantities in three adult orangutans (*Pongo pygmaeus*) ages 11, 18, and 23 (Call, 2000). In a series of discrimination experiments designed to evaluate numerousness judgments and performing operating on numbers; the orangutans were first simultaneously presented with two quantities and allowed to choose one, then in the second test they were simultaneously presented with two quantities which were then made perceptually unavailable at the time of choice, and finally in the last test they were presented with two quantities sequentially with each quantity made perceptually unavailable after presentation. The results indicated that the orangutans selected larger quantities significantly more often than smaller quantities for all three conditions. Additionally, larger differences between quantities and smaller ratios between quantities produced a higher percentage of correct trials, while total number of quantities did not affect performance. The authors concluded that the orangutans were capable of accurately estimating quantities.

In the second series of tests, Call (2000) investigated the orangutans' capability to perform mental combinations and subtractions. Orangutans were simultaneously presented with two quantities of food, one of which was added to or subtracted from, and then both quantities were made perceptually unavailable at the time of choice. In both combination and subtraction tests subjects reliably selected the larger quantity more often than the smaller. Additionally, in combination trials subjects performed better with smaller ratios between quantities rather than larger ratios reflecting a finding that paralleled Rumbaugh and colleagues (1987) and Beran (2001). In general, these results suggested that orangutans, like chimpanzees, may be capable of mental comparisons, combination, and subtractions of quantities.

The ability to perform summation has also been demonstrated in monkeys as well as apes. After tests demonstrating relative numerousness ability in four monkeys, Terrell and Thomas (1990) demonstrated the ability of monkeys to perform summation. Pairs of polygons (2 polygons on one card and 2 polygons on the other, 2 polygons on one card and 1 polygon on the other, and 1 polygon on one card and 1 polygon the other) were presented to monkeys and the authors found that one monkey was both reliable and met the criterion of 90% correct responses to the pairs of polygons for comparisons of 6:8 and 7:8 sides. The other three monkeys reliably selected the smaller number of sides for the comparison of 6:8 sides; however, they did not reach the criterion of 90% correct. Olthof and colleagues (1997) demonstrated summation in monkeys presented with two pairs of Arabic numerals with each numeral consisting of a value of zero, one, three, five, seven, or nine. Subjects received the number of food items represented by the total sum of the Arabic numeral pair they selected. Both monkeys reliably selected the larger total pair of numerals rather than the smaller total pair for all comparisons exception of 5:6; only one monkey reliably selected the larger total pair for 5:6. Analyses of their choice behavior revealed that the tendency to choose the largest pair total was not based on the avoidance of the smallest single numeral or on choosing the largest single numerals. Also, like in chimpanzees and orangutans numeral pair totals differing by small values were harder for the monkeys to discriminate between.

Aging and Numerical Ability in Humans

The definition of “number” can be stated as the only property of a set that remains invariant under substitutions of any items in the set (Dehaene, Dehaene-Lambertz, & Cohen, 1998). Because age-related changes in the numerical abilities of apes has not yet been explored, age-related changes in the human population are relevant in this discussion. Evaluation of aging and numerical ability in humans is complicated by several factors, such as differences in schooling and cohort (Geary, Bow-Thomas, Liu, & Siegler, 1996; Geary, Salthouse, Chen, & Fan, 1996; Geary, Hamson, Chen, Liu, Hoard, & Salthouse, 1997), differences in the strategies used to solve arithmetic problems (Geary & Wiley, 1991; LeFevre, Sadesky, & Bisanz, 1996; Geary & Lin, 1998), and differences in the component processes involved in performing arithmetic problems (Dehaene & Cohen, 1991; Geary, 1993; Geary & Lin, 1998). Because of complications in isolating numerical abilities in the human population, some reports indicate that older adults are slower and less accurate than younger adults in the overall performance of counting, subitizing, and in arithmetic processes of basic addition and subtraction. Alternatively, some reports indicate no difference in speed and accuracy between older and younger age groups, and some reports even indicate an older adult advantage.

Geary and Lin (1998) suggest that subitizing, magnitude comparison with smaller numbers, and simple arithmetic represent primary enumeration processes—that is cognitive competencies that emerge in the context of infant and juvenile activities. Whereas, counting, magnitude comparison with large numbers, and complex arithmetic, represent secondary processes that emerge during schooling and unnatural activities. Therefore, only age-related changes in subitizing, magnitude comparisons with small numbers, and simple arithmetic in humans are relevant to this discussion as these processes may closely parallel age-related differences in relative numerousness judgments and summation in apes. Currently, age-related differences in numerousness judgments have not been explored in humans.

Geary and Lin (1998) evaluated subitizing and magnitude comparison in humans and found that older adults performed these tasks significantly slower than the younger adults, and older adults committed the same number of errors during subitizing as young adults (errors during magnitude comparison were not calculated). Nebes, Brady, and Reynolds (1992), Kotary and Hoyer (1995), Silwinski (1997), and Watson, Maylor, and Manson (2002) also found slower subitizing speeds in older adults when compared to younger adults and no age difference in the number of errors committed during subitizing. In contrast to finding slower subitizing speeds, Trick, Enns, and Brodeur (1996) did not find statistically different subitizing speed or accuracy rates between older and younger adults. Generally, these findings indicate that older adults are slower, but not less accurate than young adults when required to subitize.

In simple subtraction tests, Geary, Frensch, and Wiley (1993) found older adults to be slower and less accurate than younger adults in solving simple subtraction problems. Geary and Lin (1998) evaluated simple subtraction in humans and found that older adults were slower, but not less accurate than were the younger adults. Birren and Botwinick (1951) and Silwinski, Buschke, Kuslansky, Senior, and Scarisbrick (1994) investigated the performance of older and younger adults in addition problems and found that older adults performed addition tasks slower, but not less accurate than younger adults. Salthouse and Kersten (1993) also found that older adults solved addition problems significantly slower than younger adults did, but in contrast they found that older adults were more accurate than were younger adults. And Geary and Wiley (1991) found similar speeds of solving addition problems between older and younger adults, but older adults were more accurate than younger. These results indicate that older adults are in general slower in performing simple arithmetic problems of addition and subtraction, but accuracy between the old and young may vary from task to task.

Objectives and Hypotheses

The present study investigated possible age-related differences in the performance of relative numerousness judgments and summation operations in Western lowland gorillas (*Gorilla gorilla gorilla*) by replicating some of the tests of Rumbaugh, Savage-Rumbaugh, and Hegel (1987). Three experiments were conducted. The first experiment, a replication of Rumbaugh et al. (1987), tested each gorilla's ability to perform relative numerousness judgments on two quantities of food items without specific prior training to do so. The second experiment, a modification of Rumbaugh et al. (1987), tested the ability of gorilla's to perform relative numerousness judgments with specific training to do so using operant conditioning with a differential reinforcement contingency. This experiment demonstrated near-asymptotic levels of performing relative numerousness judgments in the gorillas. The final and third experiment, a replication of Rumbaugh et al. (1987), tested the ability of the gorillas to perform summation operations on two pairs of food quantities without specific prior training to do so.

I expected that the gorillas would reliably select the larger quantity from the smaller quantity in the two-choice discriminations involving relative numerousness judgments with and without prior specific training (experiment 1 and 2) based on the previously discussed examinations of relative numerousness judgment in chimpanzees and orangutans by Menzel (1960), Dooley and Gill (1977), Rumbaugh et al. (1987), Pérusse and Rumbaugh (1990), Call (2000), and Beran (2001) and in monkeys by Thomas et al. (1980), Washburn and Rumbaugh (1991), Terrell and Thomas (1990), and Olthof et al. (1997). These experiments indicated that chimpanzees, orangutans, and monkeys were capable of reliably discriminating more from less with and without specific prior training to do so.

Rumbaugh et al. (1987), Rumbaugh et al. (1988), Pérusse and Rumbaugh (1990), Call (2000), and Beran (2001) in chimpanzees and orangutans and Terrell and Thomas (1990) and Olthof et al. (1997) in monkeys, demonstrated the ability to perform summation operations of quantities and/or Arabic numerals. Thus, I predicted that the gorillas would reliably perform summation operations to discriminate the larger total paired quantity from the smaller without prior specific training to do so.

With regard to age-related differences, I hypothesized that the old gorillas would perform more poorly than the young gorillas on relative numerousness judgments with and without prior specific training and on summation tasks. This hypothesis was based on the one report in humans describing older adults as less accurate in performing simple subtraction (Geary et al., 1993). And learning and discrimination impairments were found in older chimpanzees by Riopelle and Rogers (1965) in a four-choice oddity test and learning impairments were found in older chimpanzees by Bloomstrand (1986) during acquisition of a single-stimulus pair object discrimination reversal task.

I also hypothesized that the old gorillas would respond more slowly than young gorillas on relative numerousness judgment tasks with and without prior specific training and on summation tasks. This hypothesis was based on human literature which described older adults performing subitizing, magnitude comparison, simple subtraction, and addition tasks more slowly than younger adults (Geary & Lin, 1998; Nebes et al., 1992; Salthouse & Kersten, 1993; Kotary & Hoyer, 1995; Silwinski, 1997; Watson et al., 2002) and on study which found slower response times in old monkeys in comparison to young monkeys (Bachevalier et al., 1991).

The primary objective of this study was to elucidate possible age differences in the relative numerousness and summation abilities of gorillas. Since subject age typically is not included as a variate in cognition experiments in apes, the presence or absence of age-related differences in numerical abilities has major implications for future work in cognition in apes. Also, the gorillas' patterns of performance can be compared to the known performance patterns of old and young humans in numerousness and numerosity tasks. Since relative numerousness judgments are theorized as the most basic nonhuman animal number-related ability and summation as a possible precursor to basic numerical abilities in humans, the existence or nonexistence of performance differences between old and young apes generates important implications about the nature of human numerical abilities.

Secondly, relative numerousness and summation have not been previously explored in gorillas, thus the present study adds to the existing literature on numerical competence in apes and allows for phylogenetically based comparisons among apes, monkeys, and humans. And because this study utilized a sample size larger than typically used in ape learning, memory, and cognition experiments, the findings generalize more readily to the entire captive ape population. Also, the performance of the current gorilla subjects, who were relatively experimentally naïve, may be compared to the established performance levels of the previous chimpanzee and orangutan subjects each of which possessed extensive experimental histories.

CHAPTER II

METHODS

Subjects

The subjects were six old (2 male and 4 female) and five young (4 male and 1 female) adult Western lowland gorillas (*Gorilla gorilla gorilla*) housed at Zoo Atlanta in Atlanta, Georgia (Table 1). All subjects were socially housed in large indoor and outdoor enclosures. The ages of the old gorillas ranged from 38 to 43 years old (M = 40) and the ages of the young gorillas ranged from 6 to 13 years old (M = 10). The ages of subjects were measured from the starting year of this study. Gorillas 6 years and older can be considered adults and gorillas over 30 years old can be considered old (Schaller, 1965; Tarou, Bloomsmith, Hoff, Erwin, & Maple, 2000).

All subjects had received simple learning and discrimination type tasks before, also all subjects received ongoing positive reinforcement training to aid husbandry and veterinary procedures. Subjects received a regular diet of fruits, vegetables, a commercial primate diet and water was provided ad libitum. Subjects were not food or water deprived at any time during testing.

Apparatus

The apparatus (Figure 1) consisted of two rectangular plastic trays (59.7 cm x 24.1 cm x 2.5 cm). The two trays could be independently and simultaneously pushed forward into the subjects' reach or withdrawn backwards out of reach. Each plastic tray contained two circular food wells (diameter 8.9 cm and depth 2.5 cm) that were drilled into the plastic tray, one plastic rectangular stop (5.1 cm x

22.9 cm x 10.2 cm) bolted to the plastic trays, and one brass handle (length 12.7 cm) screwed into the plastic tray. The two food wells were positioned 2.5 cm from the end of the plastic tray and they were spaced 12.1 cm apart (from each wells center) and 1.6 cm from the edges of the plastic tray. The rectangular stop was positioned 17.8 cm from the end of the tray and prevented the trays from being pulled completely into the cage by the subject. The brass handle was positioned 54.6 cm from the end of the tray and 5.1 cm from the edges of the tray. The handle allowed the experimenter to pull or push the trays. The apparatus was functionally similar to that used by Rumbaugh et al. (1987).

Either grapes or occasionally Trix® cereal were used placed in the food wells of the apparatus as quantity stimuli during testing depending on the availability of grapes. Different varieties of grapes, varying in average weight and size, were used throughout the experiments. The size and weight of the grapes was not standardized, however grapes notably larger or smaller than average size within a variety of grapes were not used. Four of six different shapes of cereal pieces were used as stimuli. The four cereal pieces were shaped and colored in the following manner: a blue and red circular shape, a red circular shape, a purple circular shape, and a green and pink circular shape. Each cereal piece used as stimuli was randomly selected from all four types during testing.

Table 1. Subject name, sex, age, age class, birth year, birth type, and testing type

Age Class	Subject Name	Sex	Age	Birth Year	Birth Type	Testing Type
Y	Charlie	M	6	1996	C	1, 3
Y	Kekla	M	13	1989	C	1
Y	Kudzoo	F	8	1994	C	1, 2
Y	Stadi	M	11	1991	C	1
Y	Taz	M	13	1989	C	1
O	Banga	F	37	1965	W	4
O	Ivan	M	40	~1962	W	1, 2
O	Katoomba	F	39	1963	W	1
O	Ozoum	M	41	~1961	W	1, 2
O	Paki	F	39	1963	W	1, 3
O	Shamba	F	43	1959	W	1

Note. Y = young; O = old.
F = female; M = male.
W = wild born; C = captive born.
1 = isolated; 2 = with subordinate female(s); 3 = with adolescent; 4 = with infant.

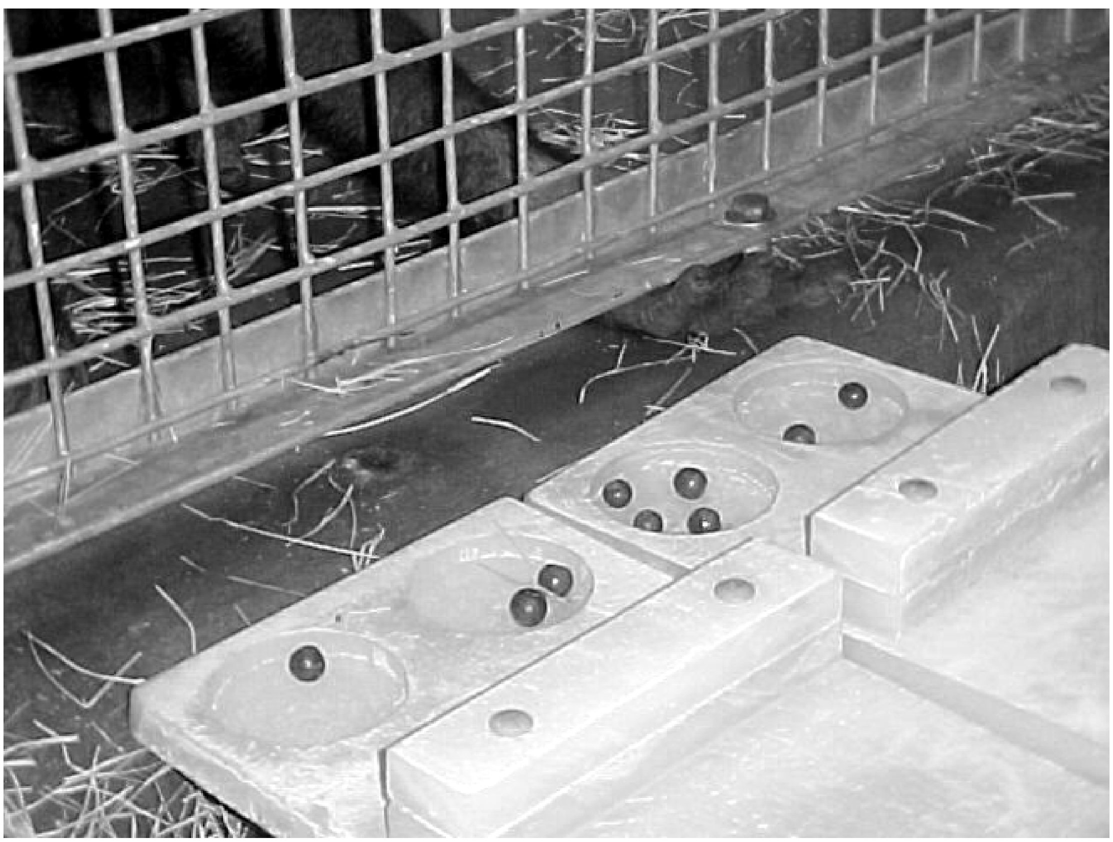


Figure 1. Testing Apparatus.

Procedure

A single experimenter collected data between June 2002 and February 2003. The experimenter wore a hat with a bill drawn low to cover the eyes or sunglasses to eliminate cueing by eye gaze. Testing occurred with subjects in their indoor cages in the afternoon, before or after subjects consumed their final meal. During testing, subjects were either isolated, with an infant, with an adolescent, or with a subordinate female(s) in their indoor cage (see Table 1). Isolation is a routine procedure during feeding and thus was not expected to disrupt performance. Infants and adolescents were not included as subjects in this study and typically sat and watched nearby as the subject was tested. When a subject was tested with a subordinate female(s) also in the cage, the subordinate female(s) did not attempt to participate and remained in the back of the cage unable to view the food wells of the apparatus.

The eleven subjects were broken into testing sets totaling one, two, three, or four subjects. Subjects of the same testing set were tested sequentially on the same day or in the same period of days. The testing sets, testing set order, and subject testing order within the testing set were determined pseudo-randomly due to the influence of husbandry considerations. Set A included subjects Kekla, Stadi, and Taz; set B Charlie, Katoomba, and Ozoum; set C Banga and Paki; set D Ivan and Shamba; and set E Kudzoo.

Testing set order occurred in the following manner for habituation and shaping and experiment 1. First, set A completed habituation and shaping followed by completing experiment 1 testing. Then, set B and C concurrently completed habituation and shaping followed by completing experiment 1 testing. And finally, set D and E concurrently completed habituation and shaping followed by completing experiment 1 testing. The testing set order for experiment 2 and 3 occurred in the following manner. First, set B completed experiment 2 and then experiment 3. Next, set C completed experiment 2 and then experiment 3. After that, set D completed experiment 2 and then experiment 3. Then, set A completed experiment 2 and then experiment 3. Finally, set E completed experiment 2 and then experiment 3.

During habituation, shaping, and testing the experimenter sat in front of each subject's cage with the apparatus placed on the floor in between the experimenter and the cage. Placed next to the experimenter was an opaque container with either grapes or occasionally cereal inside. Tray wells were baited by the experimenter placing a predetermined quantity of items into a predetermined well of each tray. Subjects could not reach the trays during baiting, but they could observe the baiting process. The experimenter prevented subjects from seeing the quantity of items selected by palming them in the hand before their removal from the opaque container until the items were dropped into a food well. After the items were dropped into a well no attempt was made to position them; they were allowed to roll or drop into random patterns within each well.

<u>Habituation and Shaping</u>

Before each subject testing set (A, B, C, D, and E) began the first experiment, individuals of that testing set were trained to choose one tray from the two trays presented. To shape subjects to touch the tray, only one tray was used at first. A quantity of food, from one to four, was placed in one well and the other well remained empty. The quantity of food and the left or right well placement was randomly determined before beginning. The single tray was then pushed forward while the experimenter gave the verbal start signal "[Subject name], choose" or "[Subject name], which one". When the subject touched the tray, it was pushed into the cage for the subject to consume.

After subjects reliably touched the single tray, the second tray was added. Identical quantities were placed in one well of each tray, such that one well remained empty on each tray. The quantities placed in the two food wells ranged from one to four. Identical quantities were used so that the conditions of experiment 1, 2, or 3 were not duplicated. Both trays were simultaneously pushed forward and at the same time the experimenter gave the verbal start signal. When subjects touched only one of the two trays, the touched tray was pushed into the cage for the subjects to consume. When subjects touched both of the trays, the experimenter simultaneously withdrew both trays and also said "No" to signal an inappropriate response. Each subject was given 20 minutes of shaping for five days. By the fifth day all subjects responded appropriately by choosing one tray from the two for all trials given.

<u>Experiment 1</u>

The conditions of experiment 1 were comparable to those used by Rumbaugh et al. (1987). A discrete quantity of food items, from zero to four, was placed into one well from each tray with the other tray well remaining empty (as in a two-choice discrimination task). The quantities were simultaneously dropped into the two wells from the experimenter's palmed hand. Trials began immediately after the wells were baited by the experimenter simultaneously pushing both trays forward into the subject's reach. The experimenter also intermittently gave the verbal start signal while pushing the trays forward. With the trays in this position, subjects could reach only the tip of each tray; they could not touch any of the food items.

The operational definition of a “choice” was any part of the hand contacting a tray (other body parts were never used by subjects to choose a tray). Very rarely did a subject choose both trays during testing. When this occurred, the experimenter immediately withdrew both trays out of reach and verbally signaled an inappropriate response. Then, without changing quantities, both trays were again pushed forward into the subject’s reach while giving the verbal start command at the same time. This procedure was repeated until the subject chose one tray.

Testing was non-corrective that is, subjects were allowed to consume the quantity of food contained by whichever tray they chose. After the subject chose a tray, the experimenter simultaneously pushed the selected tray underneath and into the subject’s cage while withdrawing the unselected tray out of reach. While subjects removed and consumed food items from the chosen tray, food items from the unselected tray were returned to the opaque container by the experimenter. Once subjects removed all food items from the chosen tray, it was withdrawn from inside the cage and a new trial was prepared.

A subject’s testing session was terminated if after the start of a trial the subject failed to respond for 10 minutes. The experimenter recorded the quantity of food items, subject choice, and subject response time. Response time was measured from the start of each trial (experimenter pushing the trays forward) until the subject chose a tray (subjects touching a tray). Response time was measured in seconds using a stopwatch by the experimenter. Each subject received approximately 20 trials for 5 days and trials were subject-paced.

The quantity of food placed in the two wells was randomly determined before testing. And placement in the left or right well of each tray was also randomly determined before testing. The term quantity comparison refers to the total of number of items in one tray versus another tray, i.e. 3:4 refers to a total of three items in one tray versus a total of four items in the other tray. Identical quantity comparisons (0:0, 1:1, 2:2, 3:3, and 4:4) were excluded from the possible combinations. The ten quantity comparisons subjects received and the ten different quantity arrangements are listed in Table 2.1. Quantity arrangements refers to the well to well placement of quantities, i.e. 0:3--0:4 indicates a zero and three food items were placed in separate wells of one tray and zero and four food items were placed separate wells of the other tray.

Experiment 2

This experiment was identical to experiment 1 except for the following procedures: First, Corrective testing was used. Subjects choosing the tray containing the smaller quantity (incorrect response) were not allowed to consume the selected quantity. Instead the experimenter simultaneously withdrawing both trays out of the subjects reach, returning both the smaller and the larger food quantities back to the opaque container, and waiting 30 seconds before preparing a new trial (inter-trial interval). Subjects selecting the tray containing the larger quantity (correct response) were allowed to consume the selected quantity. The procedure following correct responses were identical to those detailed in experiment 1.

Second, each subject received 10 minutes of testing each day until they reached an acquisition criterion of at least 80% of trials per day correct for two consecutive testing days. And since an incorrect response was followed by the ITI, the number of trials a subject received each day varied depending on the number of incorrect responses. Reaching the criterion concluded testing for experiment 2.

Table 2.1. Set of possible quantity comparisons and arrangements for experiment 1 and 2

Quantity Comparisons	Quantity Arrangements
0:1	0:0--0:1
0:2	0:0--0:2
0:3	0:0--0:3
0:4	0:0--0:4
1:2	0:1--0:2
1:3	0:1--0:3
1:4	0:1--0:4
2:3	0:2--0:3
2:4	0:2--0:4
3:4	0:3--0:4

Note. Left and right position of quantities was randomized.

Experiment 3

The conditions of experiment 3 were comparable to those used by Rumbaugh et al. (1987). Two pairs of quantities, from zero to four, were randomly assigned to both pairs of wells before testing. Within each pair of quantities, either a left or a right well location was also randomly determined before testing. The quantities were simultaneously dropped into two wells from the experimenter's palmed hand and then repeated again so that all four-wells were filled. Testing then proceeded according to the procedures detailed for experiment 1. Each subject received approximately 25 trials for ten days and trials were subject-paced and non-corrective.

Some types quantity comparisons were excluded from the set of possible comparisons. Quantity comparisons were excluded if the total of one pair equaled the total of the other pair of wells. For example, the comparisons 0:2--1:1 and 2:2--1:3 were excluded because the first pair sums to equal the second pair. Also, only one well in a pair could contain a zero value. For example, the comparison 0:0--

2:3 was excluded because the quantity five would be compared to zero. Also, only one pair of wells, but not both, could contain a zero value so that the quantities used in experiment 1 and 2 were not repeated. For example, the comparison 0:4--0:3 was excluded because it would duplicate a quantity comparison used in experiment 1 and 2. The use of all four-wells created a total of 28 quantity comparisons in 75 possible quantity arrangements (Table 2.2).

Table 2.2. Set of possible quantity comparisons and arrangements for experiment 3

Quantity Comparisons	Quantity Arrangements					
1:2	0:1--1:1					
1:3	0:1--1:2					
1:4	0:1--1:3	0:1--2:2				
1:5	0:1--1:4	0:1--2:3				
1:6	0:1--2:4	0:1--3:3				
1:7	0:1--3:4					
1:8	0:1--4:4					
2:3	0:2--1:2	1:1--1:2				
2:4	0:2--1:3	0:2--2:2	1:1--1:3	1:1--2:2		
2:5	0:2--1:4	0:2--2:3	1:1--1:4	1:1--2:3		
2:6	0:2--2:4	0:2--3:3	1:1--2:4	1:1--3:3		
2:7	0:2--3:4	1:1--3:4				
2:8	0:2--4:4	1:1--4:4				
3:4	0:3--1:3	0:3--2:2	1:2--1:3	1:2--2:2		
3:5	0:3--1:4	0:3--2:3	1:2--1:4	1:2--2:3		
3:6	0:3--2:4	0:3--3:3	1:2--2:4	1:2--3:3		
3:7	0:3--3:4	1:2--3:4				
3:8	0:3--4:4	1:2--4:4				
4:5	0:4--1:4	0:4--2:3	1:3--1:4	1:3--2:3	2:2--1:4	2:2--2:3
4:6	0:4--3:3	0:4--2:4	1:3--2:4	1:3--3:3	2:2--2:4	2:2--3:3
4:7	0:4--3:4	1:3--3:4	2:2--3:4			
4:8	0:4--4:4	1:3--4:4	2:2--4:4			
5:6	1:4--3:3	1:4--2:4	2:3--2:4	2:3--3:3		
5:7	1:4--3:4	2:3--3:4				
5:8	1:4--4:4	2:3--4:4				
6:7	2:4--3:4	3:3--3:4				
6:8	2:4--4:4	3:3--4:4				
7:8	3:4--4:4					

Note. Left or right position of quantities was randomized.

CHAPTER III

RESULTS

Following Rumbaugh et al. (1987), in experiment 1 and 2 comparisons of a quantity with zero, 0:1, 0:2, 0:3, and 0:4, were not included in the analyses as these comparisons did not reflect a meaningful choice. Thus, a total of ten quantity comparisons were statistically analyzed for experiments 1 and 2. And in experiment 3 the quantity comparison 2:8 was inadvertently not presented to subjects, making the total number of quantity comparisons 27 in 73 possible arrangements.

Three quantity variables, ratio, difference, and total were calculated from the quantity comparisons. The relationship between the quantity comparisons and the three quantity variables are presented in tabular format with the results. For experiments 1 and 2, the ratio between quantities was calculated by dividing the smaller quantity by the larger quantity, the difference between quantities by subtracting the larger quantity by the smaller quantity, and the total of quantities by adding the larger and smaller quantities together. For Experiment 3, the ratio between paired quantities was calculated by adding the pairs of quantities together and then dividing the smallest combined pair by the largest combined pair. The difference between paired quantities was calculated by adding quantity pairs together and then subtracting the largest combined pair from the smallest. The total of paired quantities was calculated by adding quantity pairs together and then adding the largest combined pair to the smallest combined pair.

Binomial tests indicated whether the frequency of correct responses was statistically greater than a chance performance level of 50% correct. Binomial tests with null hypothesis probability values of 0.50 (the level of chance) were performed to examine the frequency of selecting the larger quantity (a correct response) as opposed to the smaller. Fisher's exact tests were performed to evaluate differences in the frequency of correct responses by age group and by the quantity variables. The relation between percent of correct responses, response time, and the three quantity variables was evaluated by Pearson correlations. The Pearson chi-square test of association was used to evaluate differences in the frequency of correct responses within age groups and between individuals.

Nonparametric Kruskal-Wallis rank sum tests evaluated differences between individuals in response time (measured in seconds) and nonparametric Mann-Whitney rank sum tests determined differences between age groups in response times, and the number of trials and days required to reach the criterion. Wilcoxon signed-rank tests were performed to determine differences in response time for correct and incorrect responses. Descriptive statistics and visual inspection of trends were also used for analyses. All statistical tests were one-sided except where otherwise indicated.

Many of the statistical analyses were composed of several tests made on the same data of one individual (a family of data). For these instances a conservative approach was adopted to control family-wise (FW) error, the probability of making one or more Type I errors in a set of comparison tests. The Bonferroni procedure was applied by dividing the desired maximum FW error rate (FW*) by the number of comparisons to produce the Bonferroni adjusted alpha for any such two-tailed test. The Bonferroni adjusted alpha for one-tailed tests was obtained by divided the Bonferroni adjusted alpha calculated for two-tailed tests in half (Hayes, 1994). An alpha level of 0.05 was used as the FW* error. An alpha level of 0.01 was used as the level of significance for all other statistical tests not requiring an adjustment of alpha.

Experiment 1

Age Effects

Fisher's exact test was performed to evaluate age differences in the frequency of correct responses for all five testing days. The Fisher's exact test revealed that there was not a significant difference in the frequency of correct responses between young (65%) and old (61%) gorillas, $p = 0.108$. Also, young and old subjects did not differ in the frequency of correct responses by testing day, $p > 0.005$ in all cases. The Fisher exact test was performed with significance set at a Bonferroni adjusted alpha level of 0.005. The percent of correct responses for each age group are reported in Table 3 and graphically represented in Figure 2.

The frequency of correct responses did not differ by testing day for the young, 2 (4, $N = 333$) = 7.00, $p = 0.092$. The frequency of correct responses did not differ by testing day for the old, 2 (4, $N = 462$) = 7.57, $p = 0.109$. Chi-square tests (two-sided) were performed using a Bonferroni adjusted alpha of 0.01 to test significance. The percent of correct responses for each age group by testing day are reported in Table 3 and represented graphically in Figure 2.

Table 3. Percent of correct responses and standard error (SE) each testing day

	Testing Day					
Age Group	1	2	3	4	5	Mean
Young	64.71 (5.84)	53.13 (6.29)	61.67 (6.33)	74.67 (5.06)	69.70 (5.70)	65.17 (2.62)
Old	55.21 (5.10)	51.19 (5.49)	69.05 (5.07)	63.89 (4.64)	63.33 (5.11)	60.61 (2.28)

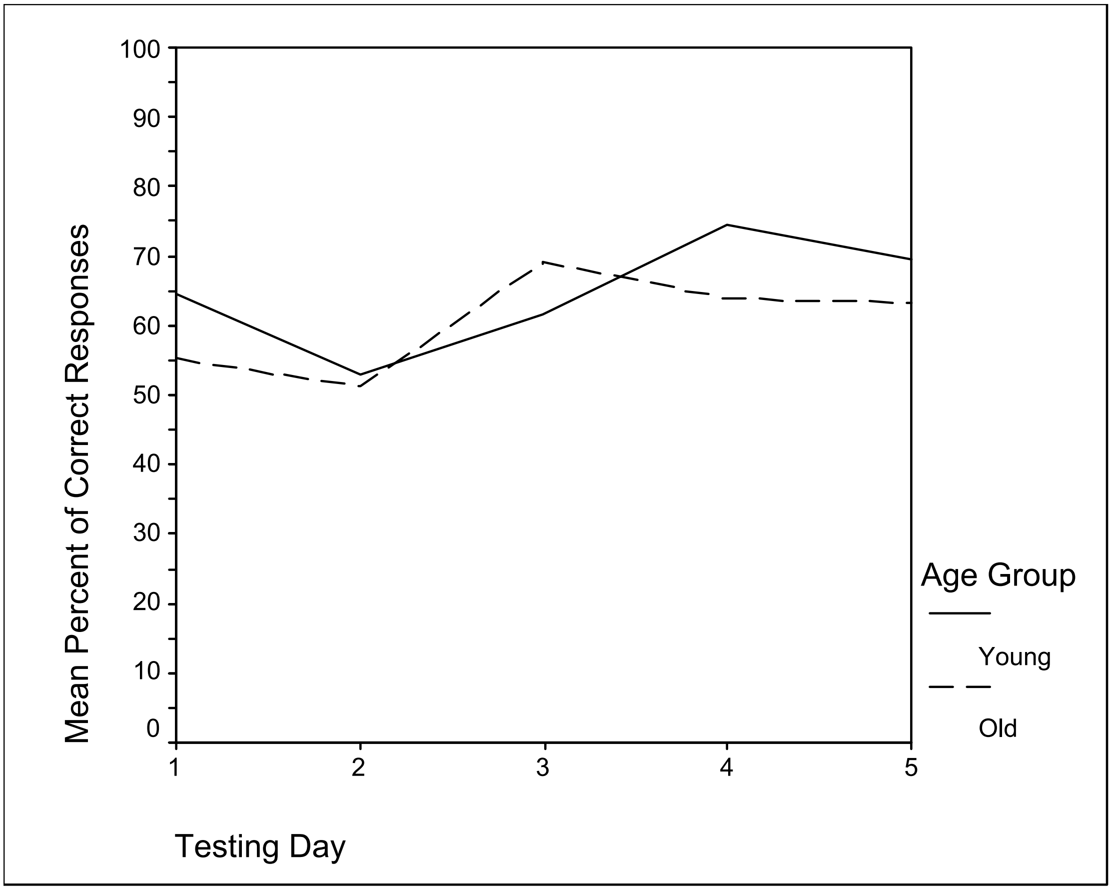


Figure 2. Percent of correct responses each testing day.

The frequency of correct responses did not differ between age groups according to the quantity variable the ratio between quantities. A Fisher's exact test with a Bonferroni adjusted alpha of 0.005 indicated that the frequency of correct responses to ratios of 0.25 (1:4), 0.33 (1:3), 0.50 (1:2 and 2:4), 0.67 (2:3), and 0.75 (3:4), were not statistically different between age groups, $p > 0.005$ (Table 4).

The frequency of correct responses did not differ between age groups according to the quantity variable the difference between quantities. A Fisher's exact test with Bonferroni adjusted alpha of 0.008 indicated that the frequency of correct responses to differences of 1 (1:2, 2:3, and 3:4), differences of 2 (1:3 and 2:4), were not significantly different between age groups, $p > 0.008$ in all cases.

Differences of 3 (1:4) were not analyzed because this difference did not contain multiple quantity comparisons, thus statistical significance was represented by the Fisher's exact test of the ratio 0.75 (Table 4).

The frequency of correct responses did not differ between age groups according to the quantity variable the total of quantities. A Fisher's exact test with Bonferroni adjusted alpha of 0.005 indicated that the frequency of correct responses to totals of 5 (1:4 and 2:3), 6 (2:4), and 3 (1:2), did not statistically differ between age groups, $p > 0.005$ in all cases. Totals of 4 (1:3) and 7 (3:4) were not analyzed because these totals did not contain multiple quantity comparisons, thus statistical significance was represented in the Fisher's exact tests by their respective ratios 0.33 and 0.75 (Table 4).

Table 4. Fisher's exact tests of frequency of correct responses to quantity variables by age group

		Young		Old		
Quantity Variables		N	Percent Correct (SE)	N	Percent Correct (SE)	p
Ratio	0.25	62	83.87 (4.71)	89	69.66 (4.90)	0.034
	0.33	49	67.35 (6.77)	92	67.39 (4.91)	0.570
	0.50	95	61.05 (5.03)	94	59.57 (5.09)	0.476
	0.67	62	56.45 (6.35)	97	53.61 (5.09)	0.426
	0.75	65	60.00 (6.124)	90	53.33 (5.29)	0.254
Difference	1	175	57.71 (3.75)	239	54.39 (3.23)	0.284
	2	96	66.67 (4.84)	134	65.67 (4.12)	0.495
Total	3	48	56.25 (7.24)	52	57.69 (6.92)	0.522
	5	124	70.16 (4.13)	186	61.29 (3.58)	0.069
	6	47	65.96 (6.99	42	61.90 (7.58)	0.430

Note. N = number of cases.

Because age differences were not found in overall performance or in performance according to the quantity variables, age groups were combined to represent all individuals as a group for subsequent analyses of the frequency of correct responses. Binomial tests performed on each quantity comparison with a Bonferroni adjusted alpha of 0.004 to test significance indicated that subjects were reliably selecting the larger quantity at quantity comparisons of 1:4 and 1:3, $\underline{p}$ = 0.001 in both cases (Figure 3). Subjects did not reliably select the larger quantity for quantity comparisons, 2:4, 1:2, 2:3, and 3:4, $\underline{p} > 0.004$ in all cases.

Pearson correlations between the percent of correct responses and each quantity variable revealed significant correlations. A Bonferroni corrected alpha of 0.008 was used as the significance level. The correlation between the percent of correct responses and the ratio between quantities was such that the percent of correct responses significantly increased as the ratio between quantities decreased, $\underline{r} = -0.25$, $\underline{p} < 0.001$. The correlation between the percent of correct responses and the difference between quantities was such that the percent of correct responses significantly increased as the difference between quantities increased $\underline{r} = 0.25$, $\underline{p} < 0.001$. The total of quantities was not significantly correlated with percent of correct responses, $\underline{r} = -0.06$, $\underline{p} = 0.140$.

Mann-Whitney tests determined that response time was not significantly different between age groups, $\underline{N} = 11$, $\underline{U} = 7.00$, $\underline{p} = 0.072$. Response time averaged 2.07 s ($\underline{SE} = 0.26$) for the young and 2.85 s ($\underline{SE} = 0.40$) for the old. Wilcoxon signed-rank tests determined that response time did not significantly differ between correct ($\underline{M} = 2.11$, $\underline{SE} = 0.27$) and incorrect ($\underline{M} = 2.02$, $\underline{SE} = 0.24$) responses for the young, $\underline{N} = 5$, $\underline{z} = -0.68$, $\underline{p} = 0.250$. Nor did response times for correct ($\underline{M} = 2.65$, $\underline{SE} = 0.37$) and incorrect ($\underline{M} = 3.28$, $\underline{SE} = 0.56$) responses differ significantly for the old, $\underline{N} = 6$, $\underline{z} = -1.36$, $\underline{p} = 0.087$. Significance was set at a Bonferroni adjusted alpha of 0.013.

Since response times did not differ between the young and old, age groups were combined to represent all individuals as a group for subsequent analyses of response time. Pearson correlations between response time and the ratio between quantities ($\underline{r} = 0.00$, $\underline{p} = 0.470$), the difference between quantities ($\underline{r} = -0.04$, $\underline{p} = 0.240$), and the total of quantities ($\underline{r} = -0.07$, $\underline{p} = 0.101$) were not statistically significant, $\underline{p}$'s > 0.008 the Bonferroni adjusted alpha. Response time for each of the quantity variables are graphed in Figure 4.1, 4.2, and 4.3. The Pearson correlation between response time and percent of correct responses was not significant, $\underline{r} = -0.18$, $\underline{p} = 0.089$.

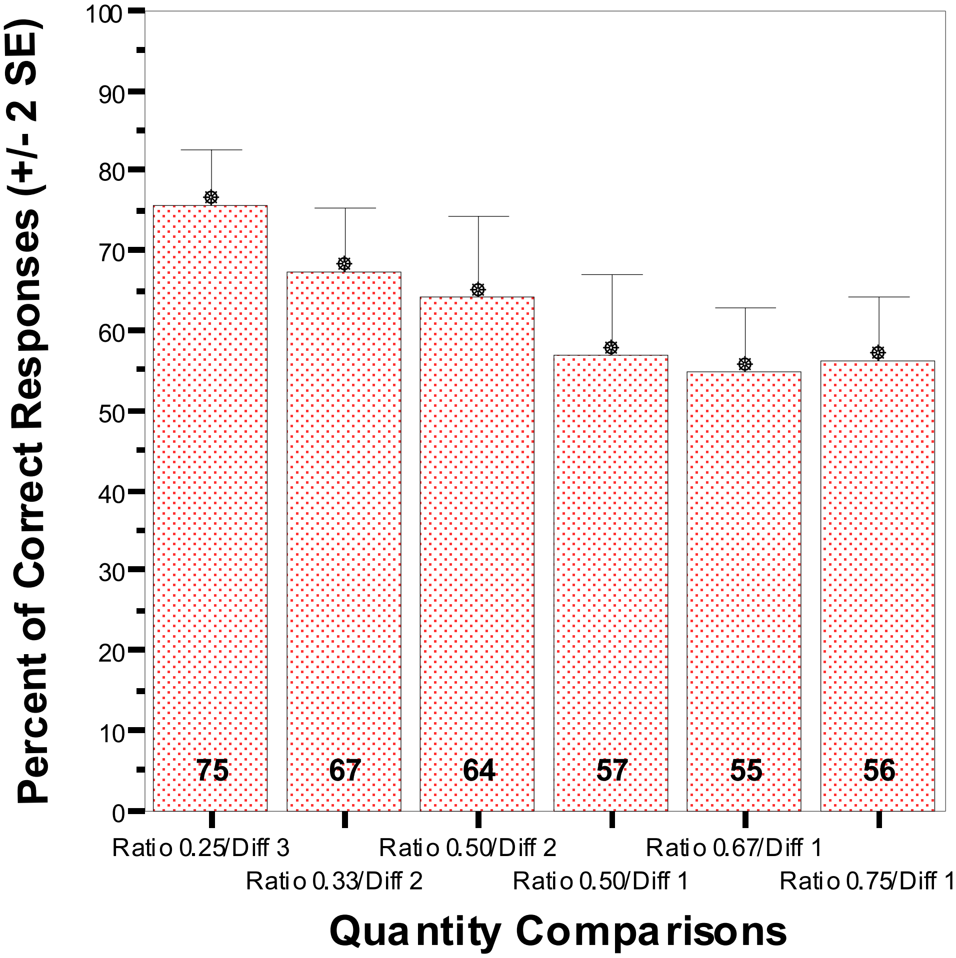


Figure 3. Percent of correct responses to quantity comparisons.

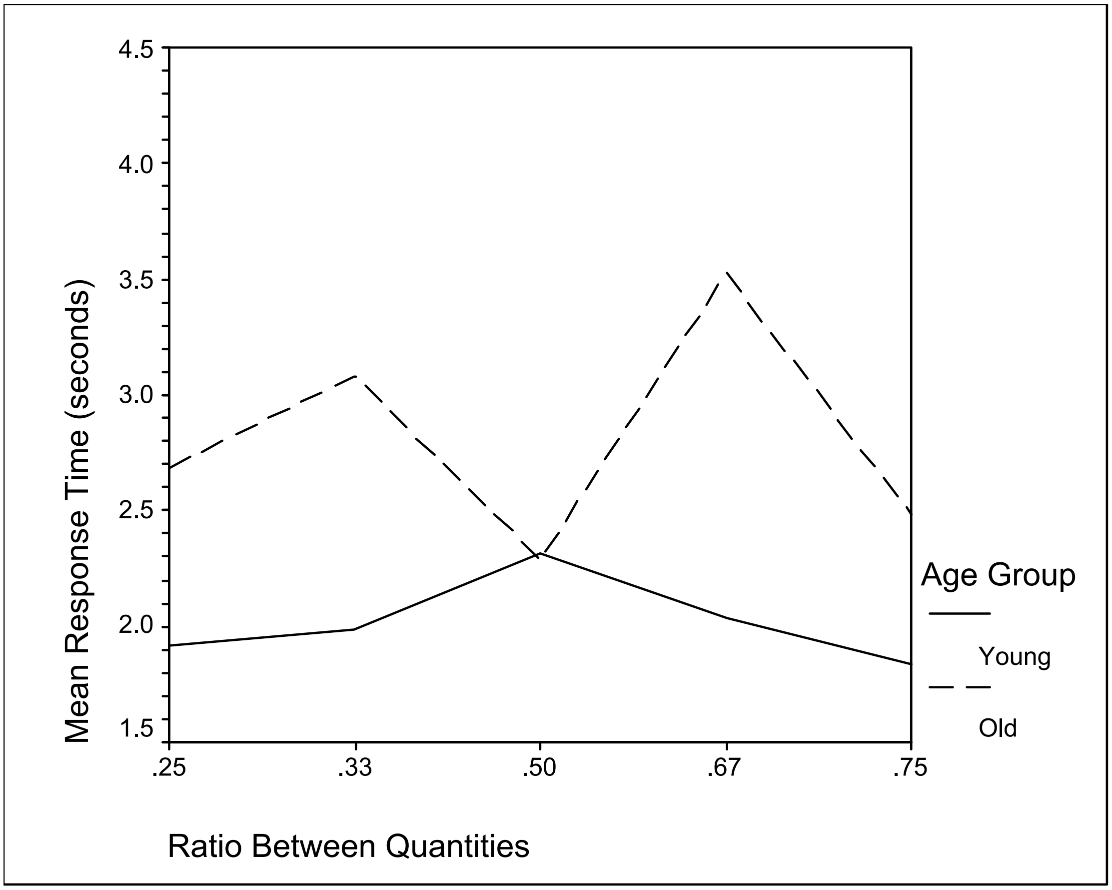


Figure 4.1. Mean response time to ratios between quantities.

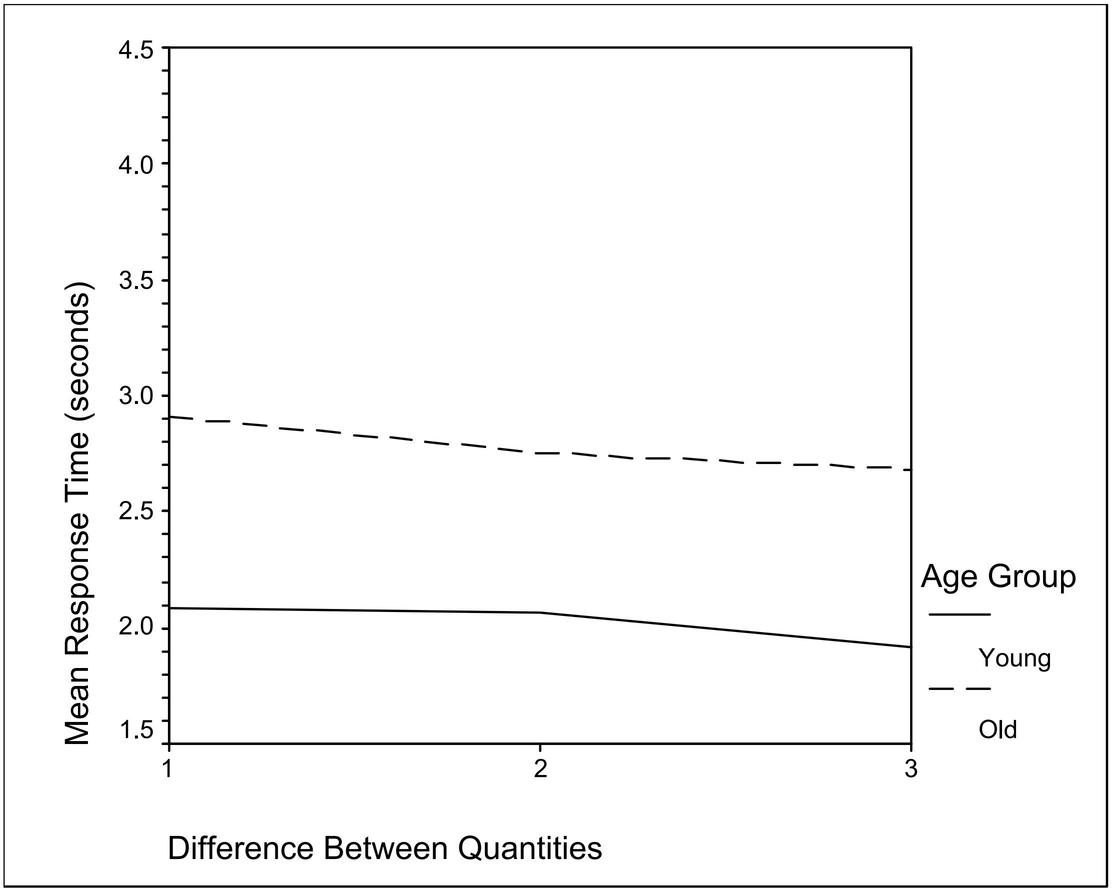


Figure 4.2. Mean response time to differences between quantities.

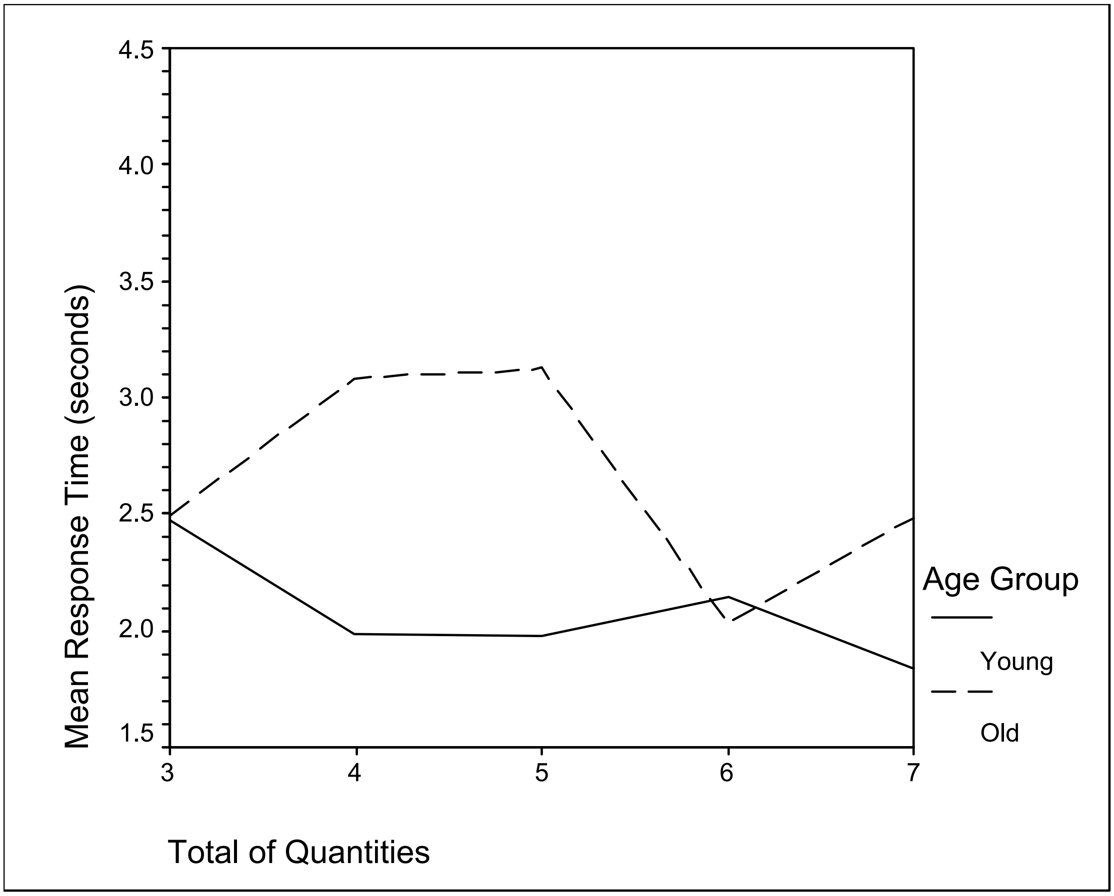


Figure 4.3. Mean response time to totals of quantities.

Individual Differences

To determine if individuals responded correctly by selecting the larger quantity more often than the smaller quantity, separate binomial tests were performed for each subject (Table 5). The binomial tests indicated that four subjects, Taz, (69%), Ivan (70%), Kudzoo (76%), and Shamba (77%), responded correctly more frequently than chance alone would allow; binomial tests, $p < 0.01$ in these cases.

Table 5. Binomial tests of the frequency of correct responses

Age Class	Subject	N	Percent Correct (SE)	p
Young	Charlie	80	60.00 (5.51)	0.047
	Kekla	57	61.40 (6.51)	0.056
	Kudzoo	80	76.25 (4.79)	0.001
	Stadi	58	56.90 (6.56)	0.179
	Taz	58	68.97 (6.13)	0.003
Old	Banga	80	53.75 (5.61)	0.288
	Ivan	74	70.27 (5.35)	0.001
	Katoomba	79	49.37 (5.66)	0.500
	Ozoum	78	61.54 (5.54)	0.027
	Paki	77	53.25 (5.72)	0.324
	Shamba	74	77.03 (4.92)	0.001

Note. N = number of cases.

Using a Bonferroni adjusted alpha level of 0.01 for significance, Chi-square tests indicated that the frequency of correct responses differed by testing day only for one subject Banga, 2 (80, N = 4) = 13.57, $p = 0.008$ (Table 6). The percent of correct responses for each subject by testing day are reported in Table 7 and represented graphically in Figure 5.1 and 5.2.

Table 6. Chi-square tests of the frequency of correct responses each testing day

Age Class	Subject	df	N	Value	p (2-tailed)
Young	Charlie	4	80	9.87	0.042
	Kekla	4	57	2.94	0.597
	Kudzoo	4	80	8.15	0.084
	Stadi	4	58	2.27	0.765
	Taz	4	58	1.47	0.872
Old	Banga	4	80	13.57	0.008
	Ivan	4	74	4.81	0.319
	Katoomba	4	79	3.66	0.466
	Ozoum	4	78	2.06	0.739
	Paki	4	77	11.00	0.025
	Shamba	4	74	2.82	0.614

Note. N= number of cases

Table 7. Percent of correct responses and standard error (SE) each testing day

Age Class	Subject	Testing Day 1	2	3	4	5	Mean
Young	Charlie	50.0 (12.91)	35.71 (13.29)	53.33 (13.33)	85.00 (8.19)	66.67 (12.60)	60.00 (5.51)
	Kekla	66.67 (14.21)	50.00 (15.08)	80.00 (13.33)	63.64 (15.21)	50.00 (15.08)	61.40 (6.51)
	Kudzoo	68.75 (11.97)	78.57 (11.38)	53.33 (13.33)	85.00 (8.19)	93.33 (6.67)	76.25 (4.79)
	Stadi	66.67 (14.21)	41.67 (14.87)	50.00 (16.67)	58.33 (14.87)	66.67 (14.21)	56.90 (6.56)
	Taz	75.00 (13.06)	58.33 (14.87)	80.00 (13.33)	66.67 (14.21)	66.67 (14.21)	68.97 (6.13)
Old	Banga	50.00 (12.91)	14.29 (9.71)	60.00 (13.09)	60.00 (11.24)	80.000 (10.69)	53.75 (5.61)
	Ivan	62.50 (12.50)	85.71 (9.71)	50.00 (16.67)	78.95 (9.61)	66.67 (12.60)	70.27 (5.35)
	Katoomba	50.00 (12.91)	42.86 (13.73)	66.67 (12.60)	52.63 (11.77)	33.33 (12.60)	49.37 (5.66)
	Ozoum	50.00 (12.91)	57.14 (13.73)	66.67 (12.60)	72.22 (10.86)	60.00 (13.09)	61.54 (5.54)
	Paki	43.75 (12.81)	21.43 (11.38)	78.57 (11.38)	55.56 (12.05)	66.67 (12.60)	53.25 (5.72)
	Shamba	75.00 (11.18)	85.71 (9.71)	86.67 (9.09)	64.29 (13.29)	73.33 (11.82)	77.03 (4.92)

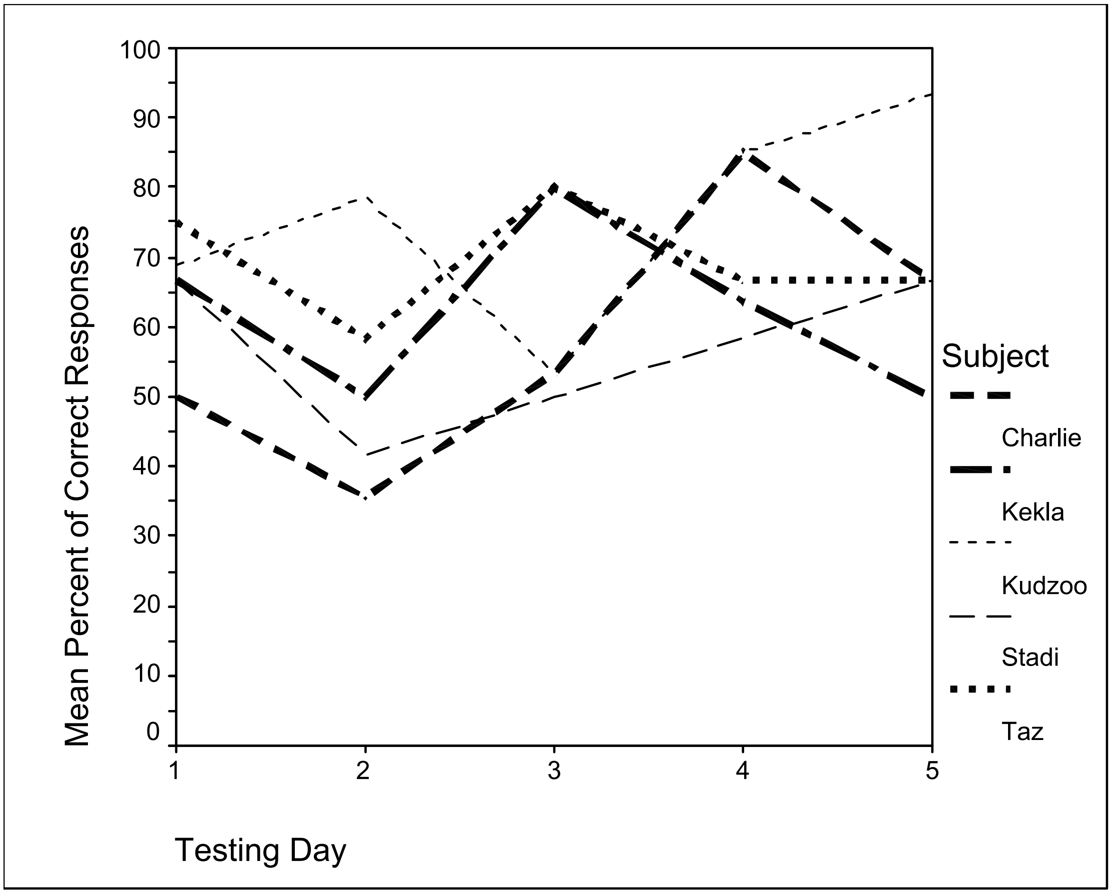


Figure 5.1. Percent of correct responses for the young each testing day.

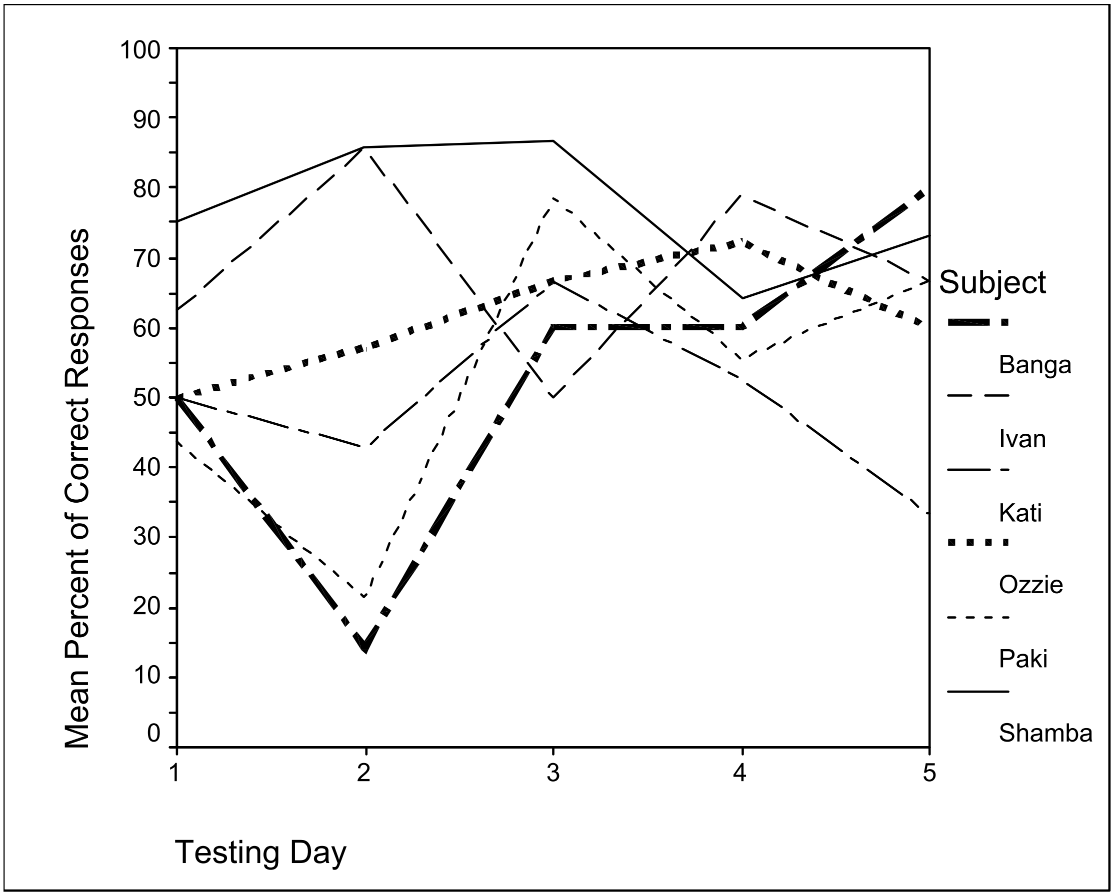


Figure 5.2. Percent of correct responses for the old each testing day.

Pearson correlations were computed to determine the correlation between percent of correct responses and each of the three quantity variables ratio, difference, and total (Table 8). The Bonferroni adjusted alpha level of 0.008 was used for significance levels. The significant relationships were such that as the difference between quantities increased the percent of correct responses increased for Kekla, Stadi, and Shamba, $\underline{r}$'s = 0.60, 0.52, and 0.51, $\underline{p} < 0.008$ respectively. Also, the percent of correct responses increased as the ratio between quantities decreased for Kekla, Stadi, and Shamba, $\underline{r}$'s = -0.48, -0.46, and -0.48, $\underline{p} < 0.008$ respectively.

Table 8. Pearson correlations between percent of correct response and the quantity variables

Age Class	Subject	Ratio	Difference	Total
Young	Charlie	-0.08	-0.05	-0.23
	Kekla	-0.48*	0.60*	0.04
	Kudzoo	-0.18	0.30	0.16
	Stadi	-0.46*	0.52*	-0.06
	Taz	-0.29	0.27	-0.06
Old	Banga	-0.04	0.10	0.13
	Ivan	-0.47	0.43	-0.20
	Katoomba	-0.38	0.30	-0.30
	Ozoum	-0.03	0.21	0.34
	Paki	-0.03	-0.22	-0.39
	Shamba	-0.48*	0.51*	-0.10

*$p < 0.008$.

Binomial tests were performed on the frequency of correct responses to evaluate individual performance differences in selecting certain quantity comparisons (Table 9). A Bonferroni adjusted alpha of 0.004 was used to determine significance. Four subjects, Kekla (100%), Kudzoo (94%), Ivan (88%), and Shamba (100%) selected the larger quantity more often than chance would allow for the comparison 1:4, $p < 0.004$. One subject, Shamba (92%) selected the larger quantity more often than chance for the 1:3 comparison, $p = 0.003$.

Table 9. Percent of correct responses to the quantity comparisons

Age Class	Subject	Quantity Comparison	Ratio	Difference	Total	N	Percent Correct (SE)	p
Young	Charlie	1:4	0.25	3	5	16	56.25 (12.81)	0.402
		1:3	0.33	2	4	16	62.50 (12.50)	0.227
		2:4	0.50	2	6	7	71.43 (18.44)	0.227
		1:2	0.50	1	3	9	66.67 (16.67)	0.254
		2:3	0.67	1	5	16	56.25 (12.80)	0.402
		3:4	0.75	1	7	16	56.25 (12.81)	0.402
	Kekla	**1:4**	**0.25**	**3**	**5**	**10**	**100.00 (0.00)**	**0.001**
		1:3	0.33	2	4	5	60.00 (24.50)	0.500
		2:4	0.50	2	6	10	63.64 (15.21)	0.274
		1:2	0.50	1	3	11	40.00 (16.33)	0.377
		2:3	0.67	1	5	10	60.00 (16.33)	0.377
		3:4	0.75	1	7	11	45.46 (15.75)	0.500
	Kudzoo	**1:4**	**0.25**	**3**	**5**	**16**	**93.75 (6.25)**	**0.001**
		1:3	0.33	2	4	16	81.25 (10.08)	0.011
		2:4	0.50	2	6	7	71.43 (18.44)	0.023
		1:2	0.50	1	3	9	44.44 (17.57)	0.500
		2:3	0.67	1	5	16	68.75 (11.97)	0.105
		3:4	0.75	1	7	16	81.25 (10.08)	0.011
	Stadi	1:4	0.25	3	5	10	90.00 (10.00)	0.011
		1:3	0.33	2	4	6	33.33 (21.08)	0.344
		2:4	0.50	2	6	11	72.73 (14.08)	0.113
		1:2	0.50	1	3	10	60.00 (16.33)	0.377

		2:3	0.67	1	5	10	40.00 (16.33)	0.377
		3:4	0.75	1	7	11	36.36 (15.21)	0.274
	Taz	1:4	0.25	3	5	10	90.00 (10.00)	0.011
		1:3	0.33	2	4	6	83.33 (16.67)	0.109
		2:4	0.50	2	6	11	54.55 (15.75)	0.500
		1:2	0.50	1	3	10	70.00 (15.28)	0.172
		2:3	0.67	1	5	10	50.00 (16.67)	0.500
		3:4	0.75	1	7	11	72.73 (14.08)	0.113
Old	Banga	1:4	0.25	3	5	16	50.00 (12.91)	0.500
		1:3	0.33	2	4	16	56.25 (12.81)	0.402
		2:4	0.50	2	6	7	85.71 (14.29)	0.063
		1:2	0.50	1	3	9	44.44 (17.57)	0.500
		2:3	0.67	1	5	16	37.50 (12.50)	0.227
		3:4	0.75	1	7	16	62.50 (12.50)	0.227
	Ivan	1:4	0.25	3	5	16	87.50 (8.54)	0.002
		1:3	0.33	2	4	16	81.25 (10.08)	0.011
		2:4	0.50	2	6	6	83.33 (16.67)	0.110
		1:2	0.50	1	3	7	57.14 (20.20)	0.500
		2:3	0.67	1	5	14	64.27 (13.29)	0.212
		3:4	0.75	1	7	15	46.67 (13.33)	0.500
	Katoomba	1:4	0.25	3	5	16	68.75 (11.97	0.105

Table 9 (cont'd).

		1:3	0.33	2	4	16	50.00 (12.91)	0.500
		2:4	0.50	2	6	6	33.33 (21.08)	0.344
		1:2	0.50	1	3	9	55.56	0.500

							(17.57)	
		2:3	0.67	1	5	16	50.00 (12.91)	0.500
		3:4	0.75	1	7	16	31.25 (11.97	0.105
	Ozoum	1:4	0.25	3	5	14	71.43 (12.53)	0.090
		1:3	0.33	2	4	16	62.50 (12.50)	0.227
		2:4	0.50	2	6	7	71.43 (18.44)	0.227
		1:2	0.50	1	3	9	44.44 (17.57)	0.500
		2:3	0.67	1	5	16	43.75 (12.81)	0.402
		3:4	0.75	1	7	16	75.00 (11.18)	0.038
	Paki	1:4	0.25	3	5	14	42.86 (13.73)	0.396
		1:3	0.33	2	4	16	68.75 (11.98)	0.105
		2:4	0.50	2	6	6	16.67 (16.67)	0.109
		1:2	0.50	1	3	9	77.78 (14.70)	0.090
		2:3	0.67	1	5	16	50.00 (12.91)	0.500
		3:4	0.75	1	7	16	50.00 (12.91)	0.500
	Shamba	**1:4**	**0.25**	**3**	**5**	**13**	**100.00 (0.00)**	**0.001**
		1:3	**0.33**	**2**	**4**	**12**	**91.67 (8.33)**	**0.003**
		2:4	0.50	2	6	10	70.00 (15.28)	0.172
		1:2	0.50	1	3	9	66.67 (16.67)	0.254
		2:3	0.67	1	5	19	73.68 (10.38)	0.032
		3:4	0.75	1	7	11	54.55 (15.75)	0.500

Note. Significant probabilities are in bold type.
N = number of cases.

Response time differed significantly between individuals according to the Kruskal-Wallis test (two-sided), 2 (10, N = 781) = 100.42, p = 0.001. Taz (1.74), Stadi (1.75), and Charlie (1.83) were the three fastest and Ivan (4.60), Kekla (3.08), and Paki (3.06) were the three slowest (Table 10). And response time did not differ between correct and incorrect responses for any individual (Table 11). The Wilcoxon signed-ranks tests were not significant, p > 0.01 in all cases.

Table 10. Kruskal-Wallis mean rank and mean response time (RT)

Age Class	Subject	N	Mean RT (SE)	Mean Rank
Young	Charlie	80	1.84 (0.23)	265.13
	Kekla	56	3.08 (0.32)	517.41
	Kudzoo	79	1.96 (0.19)	348.80
	Stadi	56	1.75 (0.12)	325.37
	Taz	57	1.74 (0.12)	315.94
Old	Banga	80	2.64 (0.32)	468.61
	Ivan	69	4.61 (0.84)	464.22
	Katoomba	79	2.94 (0.36)	445.18
	Ozoum	77	1.94 (0.14)	348.12
	Paki	74	3.06 (0.29)	486.65
	Shamba	74	1.92 (0.16)	322.91

Note. Kruskal-Wallis test is two-sided.
N = number of cases.

Table 11. Wilcoxon signed-rank tests of response time (RT) for correct and incorrect responses

Subject	Incorrect Mean RT (SE)	Correct Mean RT (SE)	z	p
Kekla	2.83(0.47)	3.06 (0.44)	-0.14	0.446
Stadi	1.47 (0.15)	1.98 (0.18)	-2.02	0.022
Taz	1.87 (0.25)	1.68 (0.21)	-0.94	0.173
Banga	3.09 (0.60)	2.14 (0.15)	-1.75	0.040
Charlie	1.90 (0.55)	1.78 (0.43)	-0.41	0.343
Katoomba	2.61 (0.29)	3.48 (0.70)	-1.48	0.069
Ozoum	2.13 (0.30)	1.86 (0.30)	-1.21	0.112
Paki	2.87 (0.32)	3.07 (0.35)	-0.14	0.446
Ivan	6.06 (1.83)	4.55 (1.07)	-0.67	0.250
Kudzoo	1.94 (0.14)	1.93 (0.27)	-0.14	0.446
Shamba	2.77 (0.75)	1.73 (0.30)	-1.21	0.112

Note. For each subject the number of cases equals five, the number of testing days.

Pearson correlations computed between response time and each of the three quantity variables, ratio, difference, and total did not indicate any significant correlations, $p > 0.008$. Bonferroni adjusted alpha of 0.008 was used for significance (Table 12). Pearson correlations computed between response time and the percent of correct responses were not significant for any individual, $p > 0.01$ (Table 12).

Table 12. Pearson correlations between response time, the quantity variables, and percent of correct responses

Age Class	Subject	Ratio	Difference	Total	Percent Correct
Young	Charlie	-0.18	0.19	-0.08	-0.17
	Kekla	-0.09	-0.09	-0.33	0.06
	Kudzoo	0.02	-0.06	-0.10	0.10
	Stadi	0.09	-0.10	-0.02	0.24
	Taz	0.12	-0.14	0.05	-0.09
Old	Banga	-0.23	0.10	-0.21	-0.36
	Ivan	0.20	-0.25	-0.10	-0.37
	Katoomba	-0.30	0.34	-0.01	0.17
	Ozoum	-0.02	-0.16	-0.21	-0.20
	Paki	0.03	0.06	0.10	-0.25
	Shamba	0.19	-0.21	-0.03	-0.37

*$p < 0.01$ for percent correct.
**$p < 0.008$ for quantity variables.

Experiment 2

Age Effects

The Fisher's exact test was performed to evaluate age differences in the frequency of correct response for all testing days required to reach the criterion (overall performance). The Fisher's exact test revealed that there was not a significant difference in the frequency of correct responses between the young (83%, $\underline{SE}$ = 2.22) and old (75%, $\underline{SE}$ = 1.48), $\underline{p}$ = 0.011.

The Fisher's exact test was performed to evaluate age differences in the frequency of correct responses on the first day of testing to reach the criterion. A Bonferroni adjusted alpha of 0.013 was used to test significance and the test revealed that the frequency of correct responses did not differ between young (75%) and old (74%) gorillas on the first day of testing, $\underline{p}$ = 0.484. The Fisher's exact test was performed to evaluate age differences in the frequency of correct responses on the last two testing days required to reach the criterion. The last two testing days are considered near-asymptotic performance levels according to the testing criterion set for experiment 2. A Bonferroni adjusted alpha of 0.013 was used to test significance. The test revealed that there was not a significant difference in the frequency of correct responses between young (88%) and old (87%) gorillas, $\underline{p}$ = 0.389.

The Mann-Whitney test determined that the number of testing days required to reach criterion for young ($\underline{M}$ = 3.60, $\underline{SE}$ = 1.12) and old ($\underline{M}$ = 4.67, $\underline{SE}$ = 1.09) did not statistically differ between age groups, $\underline{N}$ = 11, $\underline{U}$ = 10.00, $\underline{p}$ = 0.215. Also, the Mann-Whitney test determined that the number of trials needed to reach criterion were also not statistically different, $\underline{N}$ = 11, $\underline{U}$ = 11.00, $\underline{p}$ = 0.268, between young ($\underline{M}$ = 58.20, $\underline{SE}$ = 21.42) and old ($\underline{M}$ = 81.50, $\underline{SE}$ = 22.17) gorillas.

The frequency of correct responses for the last two days did not differ between age groups according to the quantity variable the ratio between quantities. A Fisher's exact test with a Bonferroni adjusted alpha of 0.005 indicated that the frequency of correct responses to ratios of 0.25 (1:4), 0.33 (1:3), 0.50 (1:2

and 2:4), 0.67 (2:3), and 0.75 (3:4), were not statistically different between age groups, $p > 0.005$ (Table 13).

Fisher's exact tests revealed that the frequency of correct responses the last two days did not statistically differ between age groups according to the quantity variable the difference between quantities. With a Bonferroni adjusted alpha of 0.008, the frequency of correct responses to differences of 1 (1:2, 2:3, and 3:4), differences of 2 (1:3 and 2:4), were not significantly different between age groups, $p > 0.008$ in all cases. Differences of 3 (1:4) were not analyzed because this difference did not contain multiple quantity comparisons, thus statistical significance was represented by the Fisher's exact test of the ratio 0.75 (Table 13).

Fisher's exact tests of the quantity variable total revealed that the frequency of correct responses the last two days did not differ between age groups according to the quantity variable the total of quantities. With a Bonferroni adjusted alpha of 0.005, the frequency of correct responses to totals of 5 (1:4 and 2:3), 6 (2:4), and 3 (1:2), did not statistically differ between age groups, $p > 0.005$ in all cases. Totals of 4 (1:3) and 7 (3:4) were not analyzed because these totals did not contain multiple quantity comparisons, thus statistical significance was represented in the Fisher's exact tests by their respective ratios 0.33 and 0.75 (Table 13).

Table 13. Fisher's exact tests of the frequency of correct responses to quantity variables by age group for the last 2 testing days

Quantity Variables		N	Young Percent Correct (SE)	Old Percent Correct (SE)	p
Ratio	0.25	78	96.77 (3.23)	97.87 (2.13)	0.640
	0.33	75	100.00 (0.00)	80.00 (6.03)	0.007
	0.50	128	86.00 (4.96)	87.18	0.524

				(3.81)	
	0.67	79	84.62 (5.85)	85.00 (5.72	0.604
	0.75	49	72.73 (9.72)	81.48 (7.62)	0.348
Difference	1	193	82.76 (4.07)	84.91 (3.49)	0.416
	2	138	92.59 (3.60)	83.33 (4.09)	0.091
Total	3	65	88.46 (6.39)	87.18 (5.42)	0.598
	5	157	90.00 (3.61)	91.95 (2.93)	0.439
	6	63	83.33 (7.77)	87.18 (5.42)	0.470

Note. N = number of cases.

Because age differences were not found for overall performance, performance the first testing day, performance the last two testing days, or for performance depending on the quantity comparisons the last two testing days, the young and old age groups were combined to represent all individuals as a group for subsequent statistical testing of the frequency of correct responses.

Binomial tests performed on the frequency of correct responses for each quantity comparison with a Bonferroni adjusted alpha of 0.004 to test significance indicated that during the last two testing days subjects as a group were reliably selecting the larger quantity at quantity comparisons of 1:4, 1:3, 2:4, 1:2, 2:3, and 3:4, $p < 0.001$ in all cases (Figure 6).

Pearson correlations were computed to determine the correlations between percent of correct responses for the last 2 testing days at criterion and each of the three quantity variables ratio, difference, and total using an adjusted Bonferroni alpha of 0.008 to test significance. Pearson correlations between the percent of correct responses for the last 2 testing days and the ratio between quantities was not statistically significant, $r = -0.21$, $p = 0.009$. The correlation between percent of correct responses and total was not significant, $r = -0.07$, $p = 0.219$. The relation between percent of correct responses and the difference between quantities was statistically significant, $r = 0.22$, $p = 0.008$.

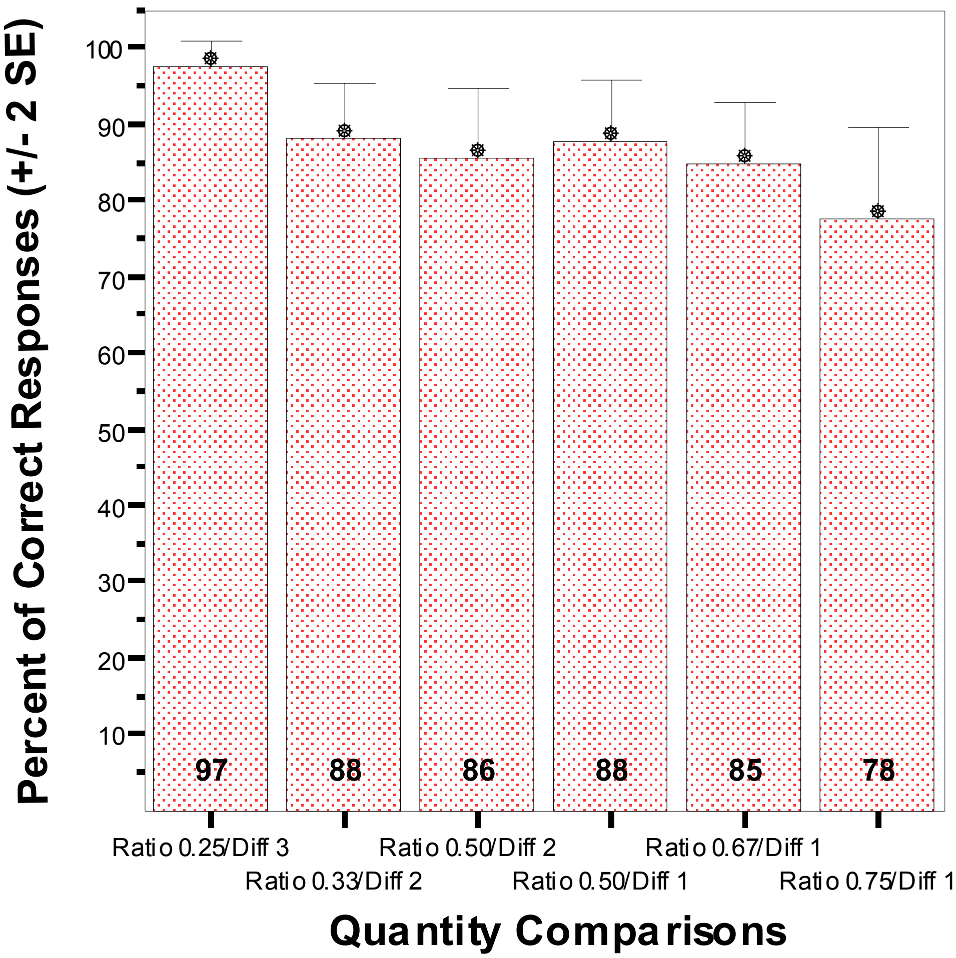


Figure 6. Percent of correct responses to the quantity comparisons for the last 2 testing days.

Mann-Whitney tests determined that response time for the first day of testing was not significantly different between the young ($\underline{M}$ = 1.96, $\underline{SE}$ = 0.40) and old ($\underline{M}$ = 2.60, $\underline{SE}$ = 0.58), $\underline{N}$ = 11, $\underline{U}$ = 11.00, $\underline{p}$ = 0.268. Mann-Whitney tests determined that response time for the last two days of testing was not significantly different between age groups, $\underline{N}$ = 11, $\underline{U}$ = 4.00, $\underline{p}$ = 0.026. Response time averaged 1.38 ($\underline{SE}$ = 0.24) for the young and 2.36 ($\underline{SE}$ = 0.29) for the old.

Wilcoxon signed-rank tests determined that response time did not significantly differ between correct ($\underline{M}$ = 1.34, $\underline{SE}$ = 0.21) and incorrect ($\underline{M}$ = 1.72, $\underline{SE}$ = 0.66) responses for the young for the last two days of testing, $\underline{N}$ = 9, $\underline{z}$ = -0.65, $\underline{p}$ = 0.258. Nor did response times for correct ($\underline{M}$ = 2.10, $\underline{SE}$ = 0.26) and incorrect ($\underline{M}$ = 3.81, SE = 1.13) responses differ significantly for the old for the last two days of testing, $\underline{N}$ = 10, $\underline{z}$ = -1.27, $\underline{p}$ = 0.102.

Pearson correlations were computed between response time for the last 2 testing days at criterion and each of the three quantity variables, ratio, difference, and total with a Bonferroni adjusted alpha of 0.008 to test significance. Correlations were not significant between the percent of correct responses and ratio for neither the young ($\underline{r}$ = -0.28, $\underline{p}$ = 0.018) nor the old ($\underline{r}$ = -0.20, $\underline{p}$ = 0.049). Correlations were also not significant between the percent of correct responses and difference for neither the young ($\underline{r}$ = 0.24, $\underline{p}$ = 0.032) nor the old ($\underline{r}$ = 0.23, $\underline{p}$ = 0.033). Also, correlations between percent of correct responses and total were not significant for the young ($\underline{r}$ = -0.11, $\underline{p}$ = 0.206) or old ($\underline{r}$ = -0.06, $\underline{p}$ = 0.309). Response times for the quantity variables are graphed in Figure 7.1, 7.2, and 7.3. The Pearson correlation between response time and percent of correct responses was not significant, for the young, $\underline{r}$ = -0.15, $\underline{p}$ = 0.340 or for the old, $\underline{r}$ = -0.05, $\underline{p}$ = 0.445.

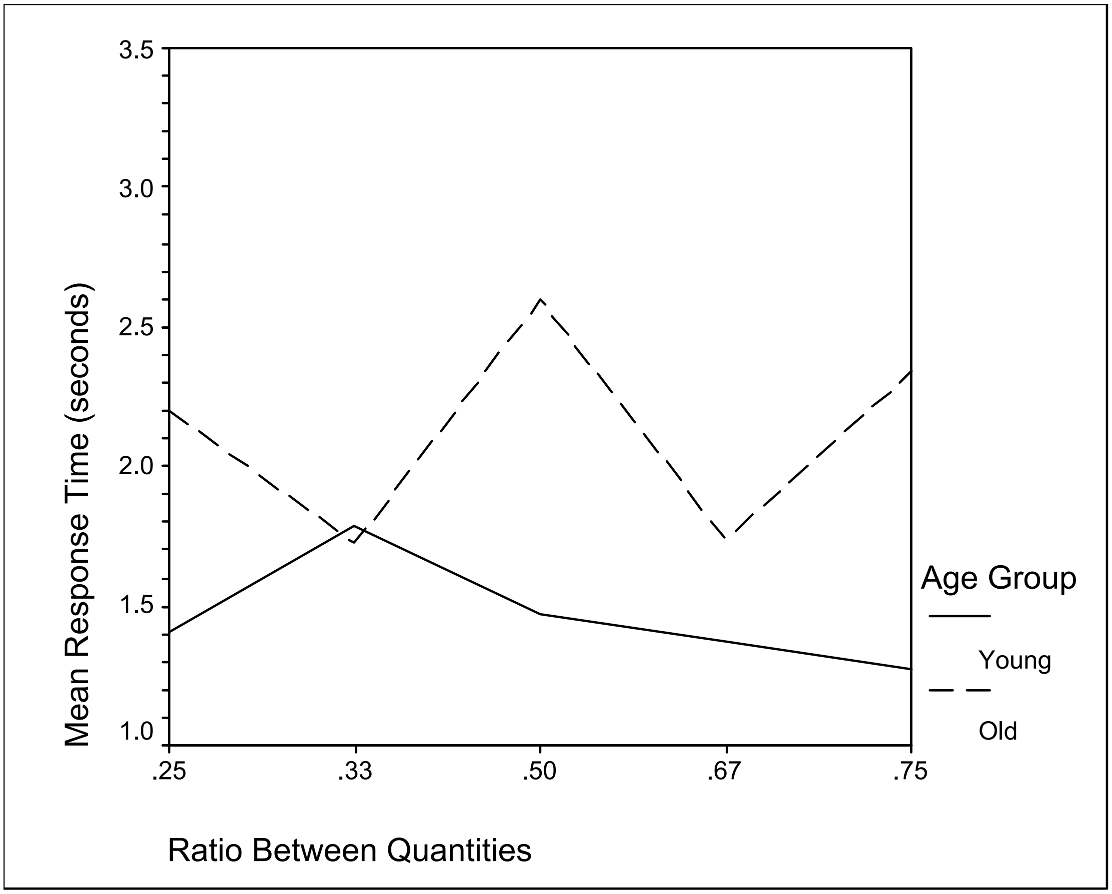
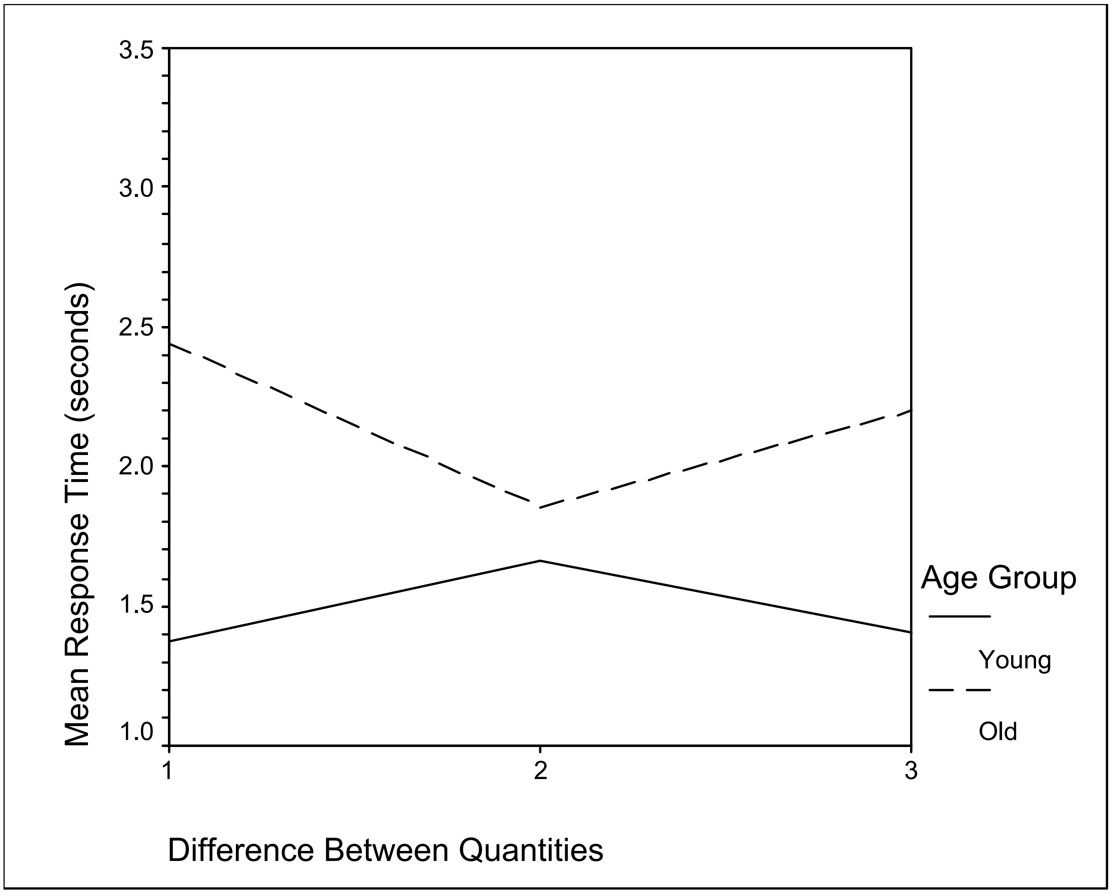


Figure 7.1. Mean response time to ratios between quantities for the last 2 testing days.



<u>Figure 7.2.</u> Mean response time to differences between quantities for the last 2 testing days.

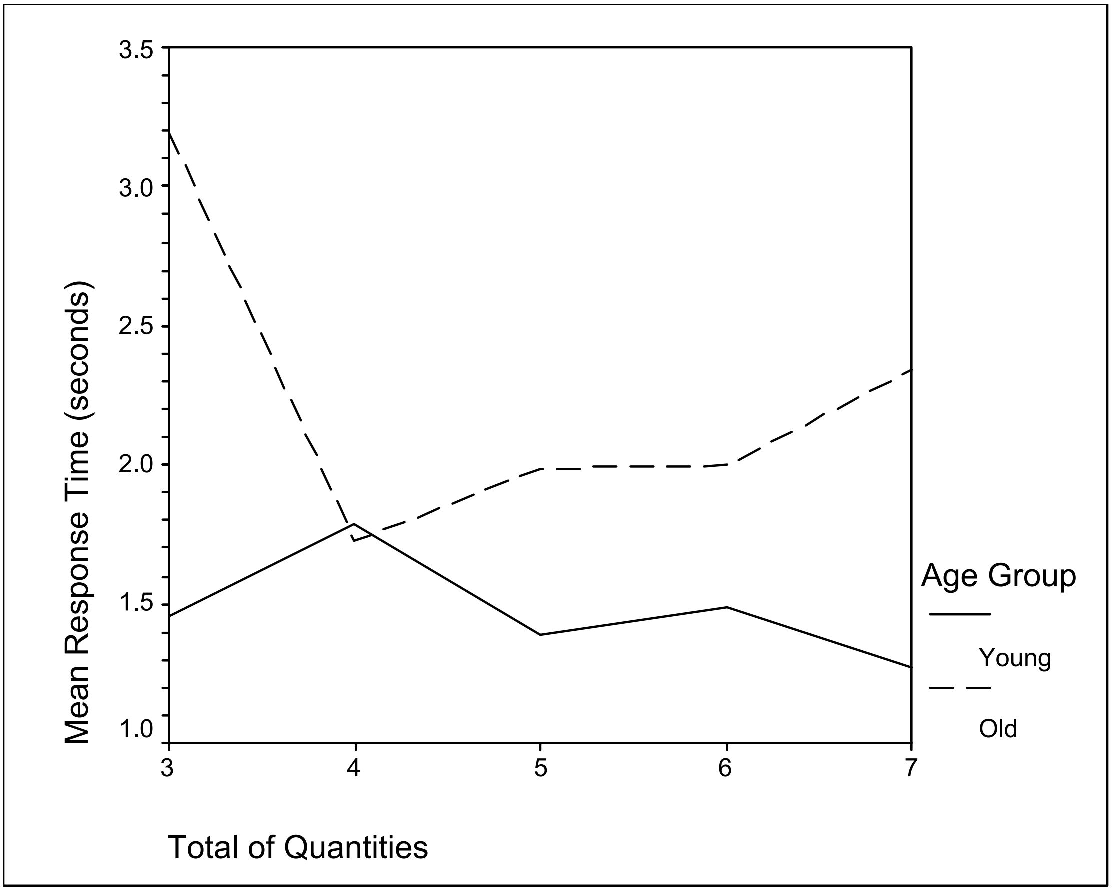


Figure 7.3. Mean response time to totals of quantities for the last 2 testing days.

Individual Differences

To determine if individuals responded correctly by selecting the larger quantity more often than the smaller quantity, separate binomial tests were performed for each subject on the frequency of correct responses for the total number of days required to reach the criterion (Table 14). The binomial tests indicated that all 11 subjects were reliably selecting the larger quantity; binomial tests, $p < 0.01$ in all cases. Binomial tests were also performed on the frequency of correct responses for the last two days of performance (Table 14). For those subjects that required only two testing days, the binomial tests were not repeated, but significance can be found by examining the probabilities listed for all testing days. The binomial tests indicated that all 11 subjects were reliably selecting the larger quantity for the last two testing days, $p < 0.01$ in all cases.

Table 14. Binomial tests of the frequency of correct responses

		All Testing Days			Last Two Testing Days		
Age Class	Subject	N	Percent Correct (SE)	p	N	Percent Correct (SE)	p
Young	Charlie	36	86.11 (5.85)	0.001			
	Kekla	141	80.14 (3.33)	0.001	48	89.58 (4.46)	0.001
	Kudzoo	53	77.36 (5.80)	0.001	34	82.35 (6.64)	0.001
	Stadi	42	92.86 (4.02)	0.001	35	97.14 (2.86)	0.001
	Taz	19	84.21 (8.60)	0.002			
Old	Banga	161	65.84 (3.75)	0.001	60	83.33 (4.85)	0.001
	Ivan	135	79.26 (3.50)	0.001	46	97.83 (2.17)	0.001
	Katoomba	37	75.67 (7.15)	0.001	25	92.00 (5.54)	0.001
	Ozoum	31	80.65 (7.21)	0.001			
	Paki	50	78.00 (5.92)	0.001	35	82.35 (6.64)	0.001
	Shamba	75	84.00 (4.26)	0.001	40	87.53 (1.64)	0.001

Note. N = number of cases.

The number of days and the number of trials each subject required to reach the criterion are listed in Table 15. Kruskal-Wallis tests indicated that both the number of testing days required to reach the criterion, 2 (10, $N = 11$) = 10.00, p = 0.440 and the number of trials required to reach the criterion, 2 (10, $N = 11$) = 10.00, p = 0.440 were not significantly different between individuals.

Table 15. Kruskal-Wallis ranks of the number of testing days and trials to criterion

		Testing Days		Trials	
Age Class	Subject	N	Rank	N	Rank
Young	Charlie	2	2.00	36	3.00
	Kekla	8	10.00	141	10.00
	Kudzoo	3	5.50	53	7.00
	Stadi	3	5.50	42	5.00
	Taz	2	2.00	19	1.00
Old	Banga	8	10.00	161	11.00
	Ivan	8	10.00	135	9.00
	Katoomba	3	5.50	37	4.00
	Ozoum	2	2.00	31	2.00
	Paki	3	5.50	50	6.00
	Shamba	4	8.00	75	8.00

Note. N = number of cases.

A chi-square test (two-sided) indicate that the frequency of correct responses on the first testing day did not differ between subjects, 2 (10, N = 152) = 12.87, p = 0.231. A chi-square test indicated that the frequency of correct responses on the second testing day was not statistically different among subjects, 2 (10, N = 217) = 7.48, p = 0.679. The percent of correct responses by each testing day the subject required to reach the criterion are listed in Table 16 and graphically represented in Figure 8.1 and 8.2.

Table 16. Percent of correct responses and standard error (SE) each testing day

		Testing Day							
Age Class	Subject	1	2	3	4	5	6	7	8
Young	Charlie	83.33 (9.04)	88.89 (7.62)						
	Kekla	60.00 (24.50)	69.23 (13.32)	72.73 (14.08)	64.71 (11.95)	88.00 (6.63)	81.82 (8.42)	88.00 (6.63)	91.30 (6.01)
	Kudzoo	68.42 (10.96)	85.71 (9.71)	80.00 (9.18)					
	Stadi	71.43 (18.44)	100.00 (0.00)	94.74 (5.26)					
	Taz	87.50 (12.50)	81.82 (12.20)						
Old	Banga	75.00 (11.18)	50.00 (15.08)	56.25 (12.81)	46.15 (14.39)	42.86 (13.73)	56.67 (9.20)	80.00 (7.43)	86.67 (6.31)
	Ivan	73.33 (11.82)	66.67 (16.67)	50.00 (18.90)	71.43 (12.53)	90.00 (6.88)	56.52 (10.57)	95.24 (4.76)	100.00 (0.00)
	Katoomba	41.67 (14.86)	100.00 (0.00)	83.33 (11.24)					
	Ozoum	81.25 (10.08)	80.00 (10.69)						
	Paki	66.67 (12.60)	80.00 (10.69)	85.00 (8.19)					
	Shamba	90.48 (6.56)	71.43 (12.53)	81.25 (10.08)	87.50 (6.90)				

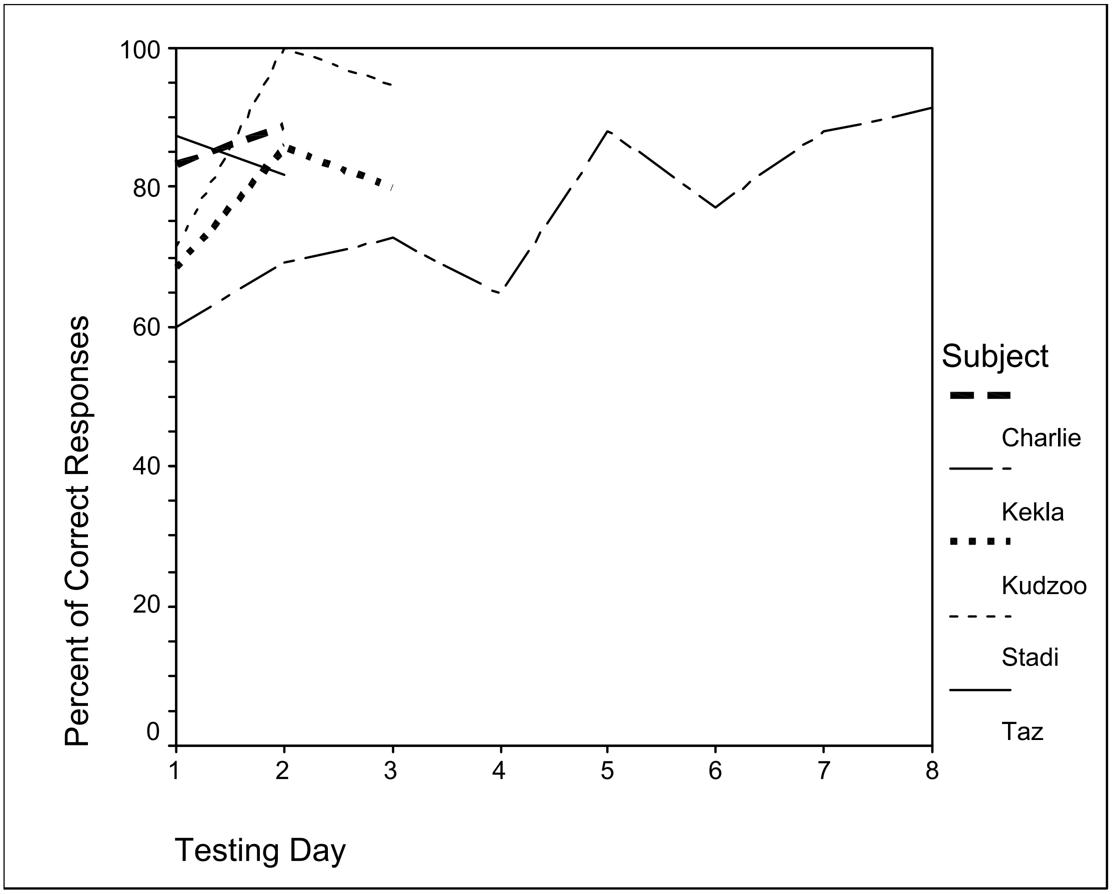


Figure 8.1. Percent of correct responses for the young each testing day.

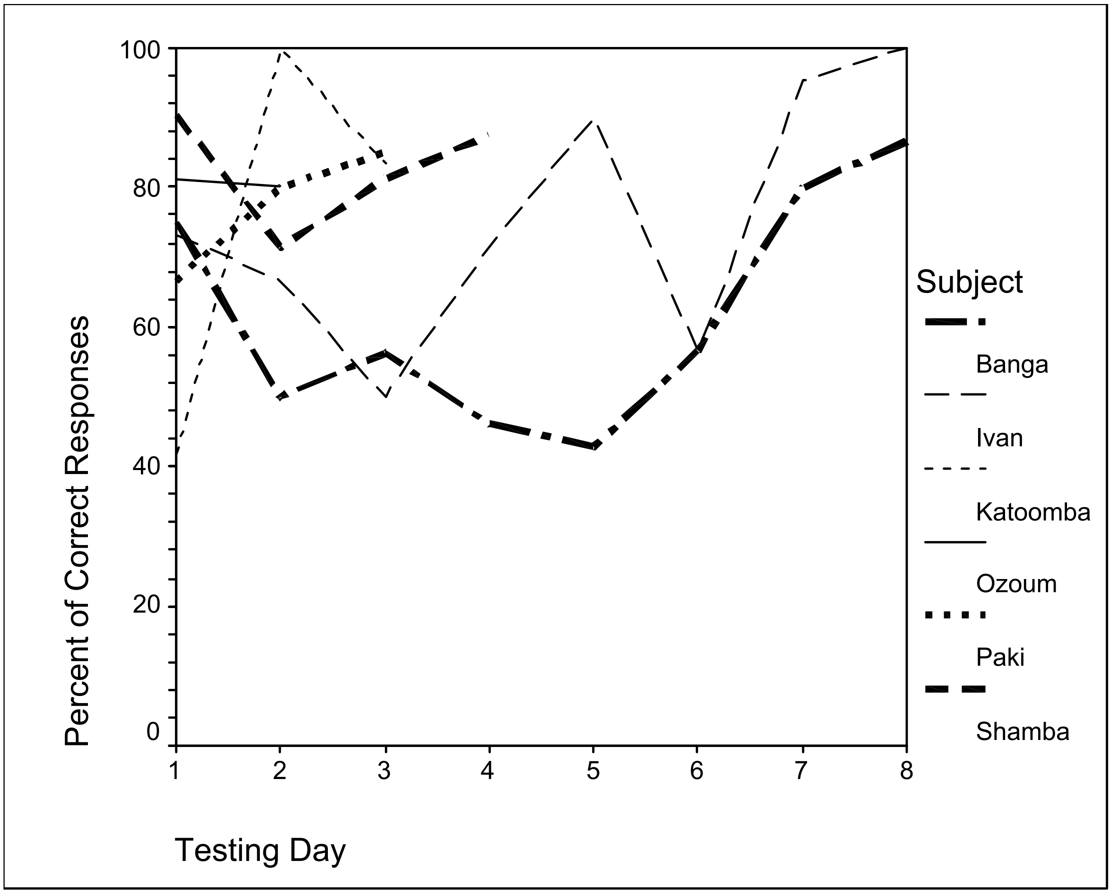


Figure 8.2. Percent of correct responses for the old each testing day.

Binomial tests were performed on the frequency of correct responses for all testing days, to evaluate individual performance differences in selecting certain quantity comparisons (Table 17). All testing days were analyzed instead of the last two testing days because the number of trials was very small for those subjects requiring four or less days to reach the criterion. A Bonferroni adjusted alpha of 0.004 was used to determine significance. The binomial tests revealed that seven subjects, Kekla (92%), Stadi (100%), Banga (83%), Paki (100%), Ivan (87%), Kudzoo (100%), and Shamba (94%) selected the larger quantity more often than chance would allow for 1:4 comparisons (ratio 0.25/difference of 3), $\underline{p} < 0.004$. The binomial tests also revealed that two subjects, Kekla (86%) and Ivan (76%), selected the larger quantity more often than chance would allow for 1:3 quantity comparisons (ratio 0.33/difference 2), $p < 0.002$. Also, the binomial tests revealed that one subject, Kekla (77%) selected the larger quantity significantly more often than chance for the quantity comparison 2:4 (ratio 0.50/difference 2), $p = 0.003$. Binomial tests indicated that two subjects, Kekla (95%) and Ivan (93%) reliably selected the larger quantity more often than chance for the quantity comparison 2:3 (ratio 0.67/difference 1), $p < 0.001$. Lastly, binomial tests indicated that Shamba (88%) reliably selected the larger quantity more than chance for quantity comparisons 3:4 (ratio 0.75/difference 1), $p = 0.002$. Taz, Charlie, Katoomba, and Ozoum did not respond correctly more often than chance for any quantity comparison.

Table 17. Binomial tests of the frequency of correct responses to quantity comparisons for all testing days

Age Class	Subject	Quantity Comparison	Ratio	Difference	N	Percent Correct Responses (SE)	p
Young	Charlie	1:4	0.25	3	7	100.00 (0.00)	0.008
		1:3	0.33	2	7	100.00 (0.00)	0.008
		2:4	0.50	2	4	50.00 (28.89)	0.500
		1:2	0.50	1	5	100.00 (0.00)	0.031
		2:3	0.67	1	8	75.00 (16.36)	0.145
		3:4	0.75	1	5	80.00 (20.00)	0.188
	Kekla	**1:4**	**0.25**	**3**	**26**	**92.31 (5.33)**	**0.001**
		1:3	**0.33**	**2**	**29**	**86.21 (5.76)**	**0.001**
		2:4	**0.50**	**2**	**30**	**76.67 (7.85)**	**0.003**
		1:2	0.50	1	13	84.62 (10.42)	0.011
		2:3	**0.67**	**1**	**20**	**95.00 (5.00)**	**0.001**
		3:4	0.75	1	23	47.83 (10.65)	0.500
	Kudzoo	**1:4**	**0.25**	**3**	**9**	**100.00 (0.00)**	**0.002**
		1:3	0.33	2	9	77.78 (14.70)	0.090
		2:4	0.50	2	5	100.00 (0.00)	0.031
		1:2	0.50	1	8	75.00 (16.37)	0.145
		2:3	0.67	1	15	60.00 (13.09)	0.304
		3:4	0.75	1	7	71.43 (18.44)	0.227
	Stadi	**1:4**	**0.25**	**3**	**8**	**100.00 (0.00)**	**0.004**
		1:3	0.33	2	7	100.00 (0.00)	0.008
		2:4	0.50	2	5	80.00 (20.00)	0.188
		1:2	0.50	1	6	100.00 (0.00)	0.016
		2:3	0.67	1	10	90.00 (10.00)	0.011
		3:4	0.75	1	6	83.33 (16.67)	0.109
	Taz	1:4	0.25	3	4	100.00 (0.00)	0.063
		1:3	0.33	2	2	100.00 (0.00)	0.250
		2:4	0.50	2	3	100.00 (0.00)	0.125
		1:2	0.50	1	4	75.00 (25.00)	0.313
		2:3	0.67	1	2	100.00 (0.00)	0.250
		3:4	0.75	1	4	50.00 (28.87)	0.500
Old	Banga	**1:4**	**0.25**	**3**	**35**	**82.86 (6.46)**	**0.001**
		1:3	0.33	2	28	71.43 (8.69)	0.018
		2:4	0.50	2	41	56.10 (7.85)	0.266
		1:2	0.50	1	23	78.26 (8.79)	0.005
		2:3	0.67	1	19	68.42 (10.96)	0.084
		3:4	0.75	1	15	20.00 (10.69)	0.018
	Ivan	**1:4**	**0.25**	**3**	**31**	**87.10 (6.12)**	**0.001**
		1:3	**0.33**	**2**	**33**	**75.76 (7.58)**	**0.002**
		2:4	0.50	2	28	75.00 (8.33)	0.007
		1:2	0.50	1	12	75.00 (13.06)	0.073
		2:3	**0.67**	**1**	**15**	**93.33 (6.67)**	**0.001**

		3:4	0.75	1	16	68.75 (11.97)	0.105

Table 17 (cont'd).

	Katoomba	1:4	0.25	3	7	57.14 (20.20)	0.500
		1:3	0.33	2	7	85.71 (14.29)	0.063
		2:4	0.50	2	4	50.00 (28.87)	0.500
		1:2	0.50	1	6	83.33 (16.67)	0.109
		2:3	0.67	1	7	85.71 (14.29)	0.063
		3:4	0.75	1	6	83.33 (16.67)	0.109
	Ozoum	1:4	0.25	3	7	100.00 (0.00)	0.008
		1:3	0.33	2	6	66.67 (21.08)	0.344
		2:4	0.50	2	3	100.00 (0.00)	0.125
		1:2	0.50	1	4	75.00 (25.00)	0.313
		2:3	0.67	1	6	83.33 (16.67)	0.109
		3:4	0.75	1	5	60.00 (24.50)	0.500
	Paki	**1:4**	**0.25**	**3**	**9**	**100.00 (0.00)**	**0.002**
		1:3	0.33	2	10	80.00 (13.33)	0.055
		2:4	0.50	2	5	80.00 (20.00)	0.188
		1:2	0.50	1	7	100.00 (0.00)	0.008
		2:3	0.67	1	12	66.67 (14.21)	0.194
		3:4	0.75	1	7	42.86 (20.20)	0.500
	Shamba	**1:4**	**0.25**	**3**	**17**	**94.12 (5.88)**	**0.001**
		1:3	0.33	2	14	78.57 (11.38)	0.029
		2:4	0.50	2	8	87.50 (12.50)	0.035
		1:2	0.50	1	8	87.50 (12.50)	0.035
		2:3	0.67	1	12	66.67 (14.21)	0.194
		3:4	**0.75**	**1**	**16**	**87.50 (8.54)**	**0.002**

Note. Significant probabilities are in bold type.
N = number of cases.

Pearson correlations were computed to determine the correlation between percent of correct responses for the last 2 testing days at criterion and each of the three quantity variables ratio, difference, and total (Table 18). A Bonferroni adjusted alpha level of 0.008 was used for significance level. Correlations between percent of correct responses for the last 2 days of testing at criterion and each of the three quantity variables were not significant for any subject, $p > 0.008$ in all cases.

Table 18. Pearson correlations between percent of correct responses and the quantity variables for the last 2 testing days

Age Class	Subject	Ratio	Difference	Total
Young	Charlie	-0.41	0.13	-0.55
	Kekla	-0.32	0.00	-0.70
	Kudzoo	-0.45	0.60	0.16
	Stadi	-0.29	0.27	0.00
	Taz	-0.39	0.37	-0.19
Old	Banga	0.03	0.09	0.13
	Ivan	0.28	-0.12	0.22
	Katoomba	0.00	0.00	0.00
	Ozoum	-0.30	0.41	0.00
	Paki	-0.43	0.24	-0.49
	Shamba	-0.03	0.37	0.54

*$p < 0.008$.

Response time (measured in seconds) for the last 2 days of testing differed significantly between individuals according to the Kruskal-Wallis test (2-sided), 2 (10, $N = 399$) = 156.52, $p = 0.001$. Stadi (0.67), Charlie (0.95), and Kudzoo (1.18) responded the three fastest and Katoomba (3.95), Ozoum (2.78), and Kekla (2.70) the three slowest. The Kruskal-Wallis mean ranks are listed in Table 19.

Table 19. Kruskal-Wallis mean rank and mean response time (RT) for the last 2 testing days

Age Class	Subject	N	Mean RT (SE)	Mean Rank
Young	Charlie	35	0.95 (0.09)	108.76
	Kekla	46	2.70 (0.44)	231.42
	Kudzoo	33	1.18 (0.07)	167.32
	Stadi	35	0.67 (0.06)	46.16
	Taz	19	1.40 (0.20)	175.89
Old	Banga	56	1.85 (0.43)	184.76
	Ivan	46	1.94 (0.13)	268.64
	Katoomba	25	3.95 (0.61)	337.22
	Ozoum	31	2.78 (0.99)	270.81
	Paki	33	2.02 (0.56)	205.39
	Shamba	40	1.48 (0.10)	214.04

Note. Kruskal-Wallis test was two-sided.
N = number of cases.

Pearson correlations computed for the last two testing days between response time and each of the three quantity variables, ratio, difference, and total did not indicate significant linear relations, $p > 0.008$ for all individuals. A Bonferroni adjusted alpha of 0.008 was used for significance (Table 20). Pearson correlations between response time and the percent of correct responses were significant for one old gorilla, Paki, $r = 0.87$, $p = 0.001$. Response time and percent of correct responses were not correlated for any other individual, $p > 0.01$ (Table 20).

Table 20. Pearson correlations between response time, quantity variables, and the percent of correct responses

Age Class	Subject	Ratio	Difference	Total	Percent Correct
Young	Charlie	0.00	0.25	0.41	0.09
	Kekla	-0.12	0.00	-0.17	-0.06
	Kudzoo	0.44	-0.60	-0.14	-0.57
	Stadi	0.13	-0.17	-.12	0.03
	Taz	0.23	-0.47	-0.45	-0.01
Old	Banga	0.20	-0.33	-0.32	0.38
	Ivan	0.13	-0.25	-0.21	0.14
	Katoomba	-0.46	0.63	0.11	-0.20
	Ozoum	-0.05	-0.22	-0.49	-0.34
	Paki	0.41	-0.25	0.44	0.87*
	Shamba	-0.18	-0.10	-0.50	-0.15

*$\underline{p} < 0.05$.
**$\underline{p} < 0.008$.

Response times for correct and incorrect responses on the last two testing days did not statistically differed for any subject (Table 21). Wilcoxon signed-ranks tests with significance set at a Bonferroni adjusted alpha of 0.013 were not significant, $\underline{p} > 0.013$ in all cases.

Table 21. Wilcoxon signed-rank tests of response time (RT) between correct and incorrect responses

Age Class	Subject	RT for Incorrect Response (SE)	RT for Correct Response (SE)	$\underline{z}$	$\underline{p}$
Young	Charlie	0.84 (0.05)	0.98 (0.11)	-0.45	0.327
	Kekla	4.33 (2.62)	2.42 (0.12)	-0.45	0.327
	Kudzoo	1.28 (0.21)	1.14 (0.04)	-0.45	0.327
	Stadi	0.54 (0.00)	0.67 (0.04)		
	Taz	1.03 (0.19)	1.49 (0.27)	-1.34	0.090
Old	Banga	1.40 (0.16)	1.99 (0.80)	-0.45	0.327
	Ivan	2.71 (0.00)	1.94 (0.02)		
	Katoomba	5.80 (0.00)	3.81 (0.20)		
	Ozoum	6.85 (4.90)	1.80 (0.12)	-1.34	0.090
	Paki	5.01 (0.00)	1.37 (0.04)	-1.34	0.090
	Shamba	2.83 (0.70)	1.74 (0.19)	-0.45	0.327

Note. Wilcoxon tests not performed when insufficient $\underline{N}$, the number of cases. $\underline{N}$ = 2 for each individual.

Experiment 3

Age Effects

Fisher's exact test was performed on the frequency of correct responses between young and old subjects for all testing days. The Fisher's exact test was significant, $p < 0.001$, the frequency of correct responses was different between the young (77%, SE = 1.19) and the old (68%, SE = 1.25).

The frequency of correct responses did not differ between age groups according to the quantity variable the ratio between quantities, $p > 0.001$ in all cases (Table 22). Fisher's exact tests with a Bonferroni adjusted alpha of 0.001 were used to test significance (Figure 9.1 and 9.2). The frequency of correct responses differed between age groups according to the quantity variable the difference between quantities. Fisher's exact tests with a Bonferroni adjusted alpha of 0.004 to test significance, indicated that the frequency of correct responses to differences of 1 (1:2, 2:3, 3:4, 4:5, 5:6, 6:7, and 7:8), 2 (1:3, 2:4, 3:5, 4:6, 5:7, and 6:8), and 5 (1:6, 2:7, and 3:8) was significantly higher for the young than the old, $p < 0.004$. Differences of 6 (1:7) and 7 (1:8) were not analyzed because this difference did not contain multiple quantity comparisons, thus statistical significance was represented by the Fisher's exact tests of the respective ratios 0.143 and 0.125 (Table 22 and Figure 10.1 and 10.2).

The frequency of correct responses differed between age groups according to the quantity variable the total of quantities. Fisher's exact tests indicated with a Bonferroni adjusted alpha of 0.002 to test significance indicated that the frequency of correct responses to totals of 9 (1:8, 2:7, 3:6, and 4:5) was significantly higher for the young than the old, $p < 0.001$. Totals of 15 (7:8) were not analyzed because this total did not contain multiple quantity comparisons, thus statistical significance was represented by the Fisher's exact tests of the ratio 0.875 (Table 22 and Figure 11.1 and 11.2).

Table 22. Fisher's exact tests of the frequency of correct responses to the quantity comparisons by age group

Quantity Variables		N	Young Percent Correct (SE)	Old Percent Correct (SE)	p
Ratio	0.125	88	100.00 (0.00)	95.65 (3.04)	0.270
	0.143	78	97.22 (2.78)	92.86 (4.02)	0.368
	0.167	85	97.50 (2.50)	88.89 (4.74)	0.130
	0.200	96	95.45 (3.18)	90.38 (4.13)	0.293
	0.250	95	93.88 (3.46)	84.78 (5.35)	0.134
	0.286	99	95.74 (2.98)	88.46 (4.47)	0.170
	0.333	94	89.47 (4.10)	72.97 (7.40)	0.037
	0.375	88	91.11 (4.29)	72.09 (6.92)	0.020
	0.400	98	91.84 (3.95)	77.55 (6.02)	0.045
	0.429	103	84.78 (5.35)	77.19 (5.61)	0.238
	0.500	152	78.38 (4.82)	73.08 (5.06)	0.284
	0.571	125	78.57 (5.53)	68.12 (5.65)	0.135
	0.600	124	74.51 (6.16)	65.75 (5.59)	0.200
	0.625	138	62.12 (6.02)	62.50 (5.75)	0.551
	0.667	130	76.56 (5.34)	63.64 (5.97)	0.078
	0.714	156	67.12 (5.54)	57.83 (5.45)	0.152
	0.750	141	68.18 (5.78)	64.00 (5.58)	0.366
	0.800	242	67.37 (4.84)	48.98 (4.14)	0.003
	0.833	236	60.00 (4.92)	50.74 (4.30)	0.100
	0.857	132	63.89 (5.70)	50.00 (6.51)	0.076
	0.875	136	61.43 (5.86)	63.64 (5.97)	0.465
Difference	**1**	**982**	**64.98 (2.24)**	**54.55 (2.17)**	**0.001**
	2	**493**	**77.02 (2.75)**	**64.73 (2.98)**	**0.002**
	3	494	79.58 (2.61)	72.44 (2.81)	0.040
	4	229	90.29 (2.93)	82.54 (3.40)	0.067
	5	**272**	**94.70 (1.96)**	**83.57 (3.14)**	**0.003**
Total	3	26	66.67 (11.43)	62.50 (18.30)	0.587
	4	75	87.76 (4.73)	69.23 (9.23)	0.052
	5	187	84.85 (3.62)	76.14 (4.57)	0.093
	6	173	92.00 (3.15)	81.63 (3.93)	0.039
	7	301	84.06 (3.13)	73.62 (3.46)	0.019
	8	221	85.26 (3.66)	76.19 (3.81)	0.065
	9	**467**	**81.37 (2.73)**	**67.30 (2.90)**	**0.001**
	10	141	83.33 (4.85)	71.60 (5.04)	0.076
	11	449	72.14 (3.17)	59.27 (3.13)	0.003
	12	167	67.95 (5.32)	58.43 (5.25)	0.133
	13	270	63.04 (4.12)	56.82 (4.33)	0.179
	14	23	76.47 (10.61)	100.00 (0.00)	0.269

Note. Significant probabilities are in bold type.
N = number of cases.

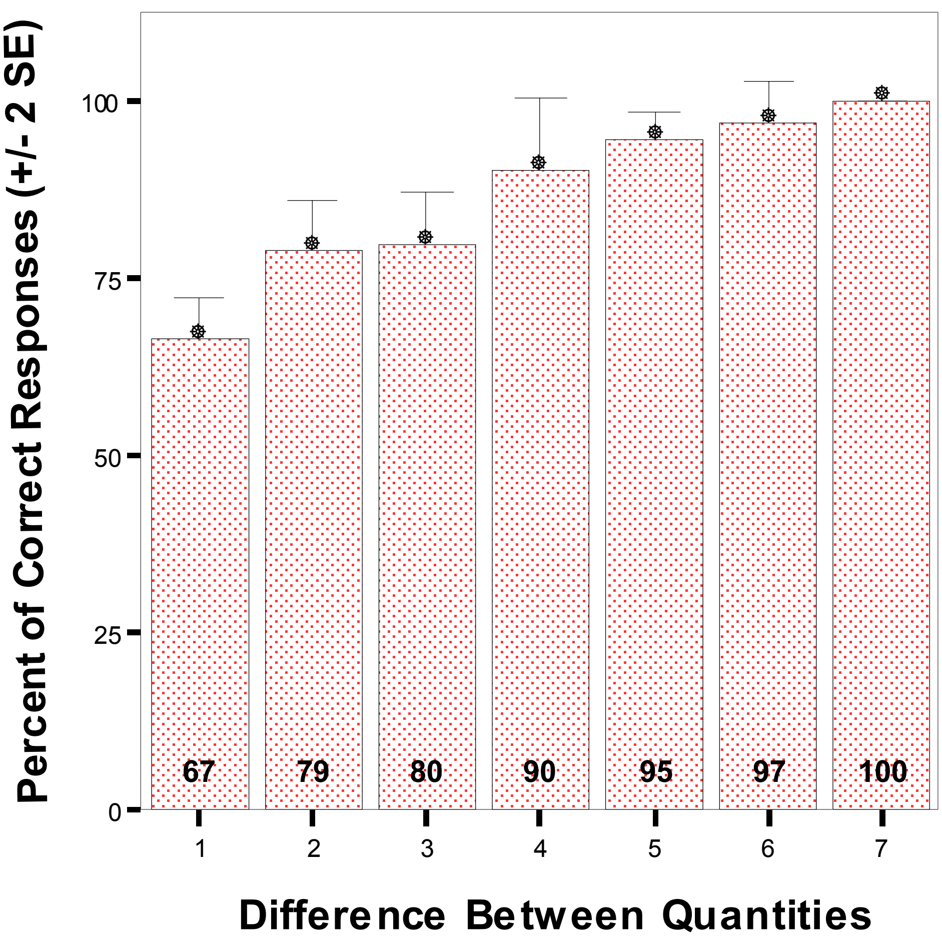


Figure 9.1. Percent of correct responses to differences between quantities for the young.

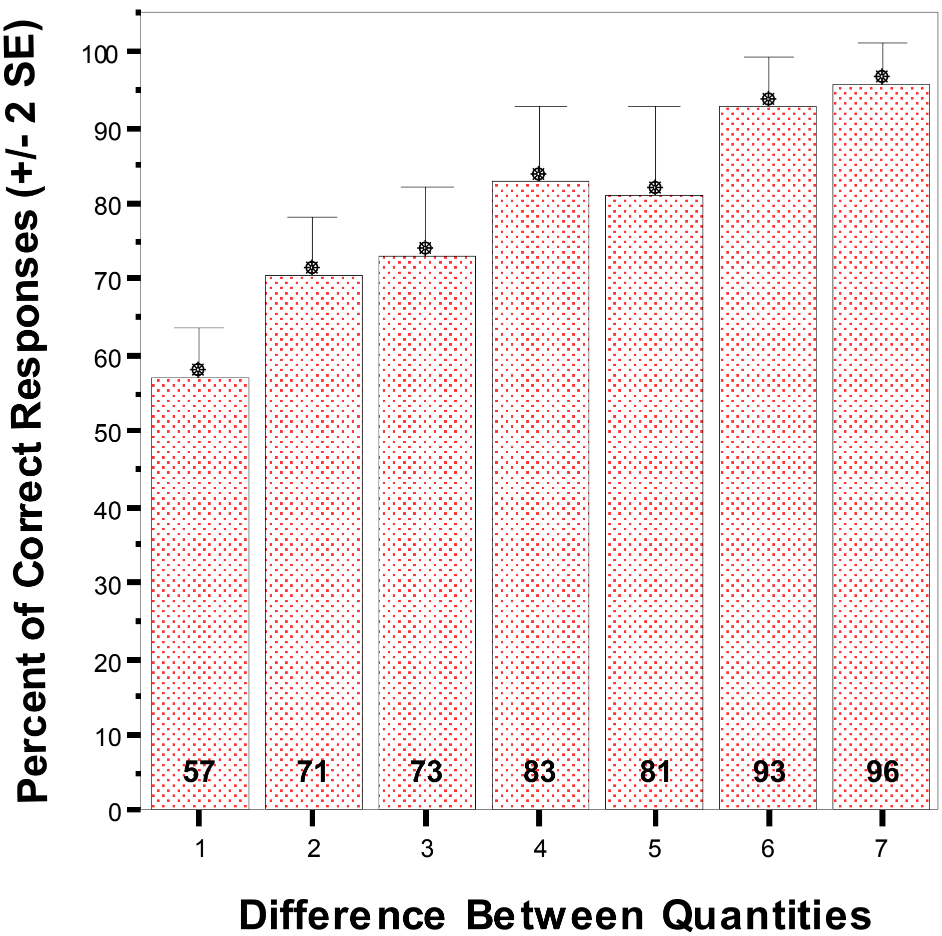


Figure 9.2. Percent of correct responses to differences between quantities for the old.

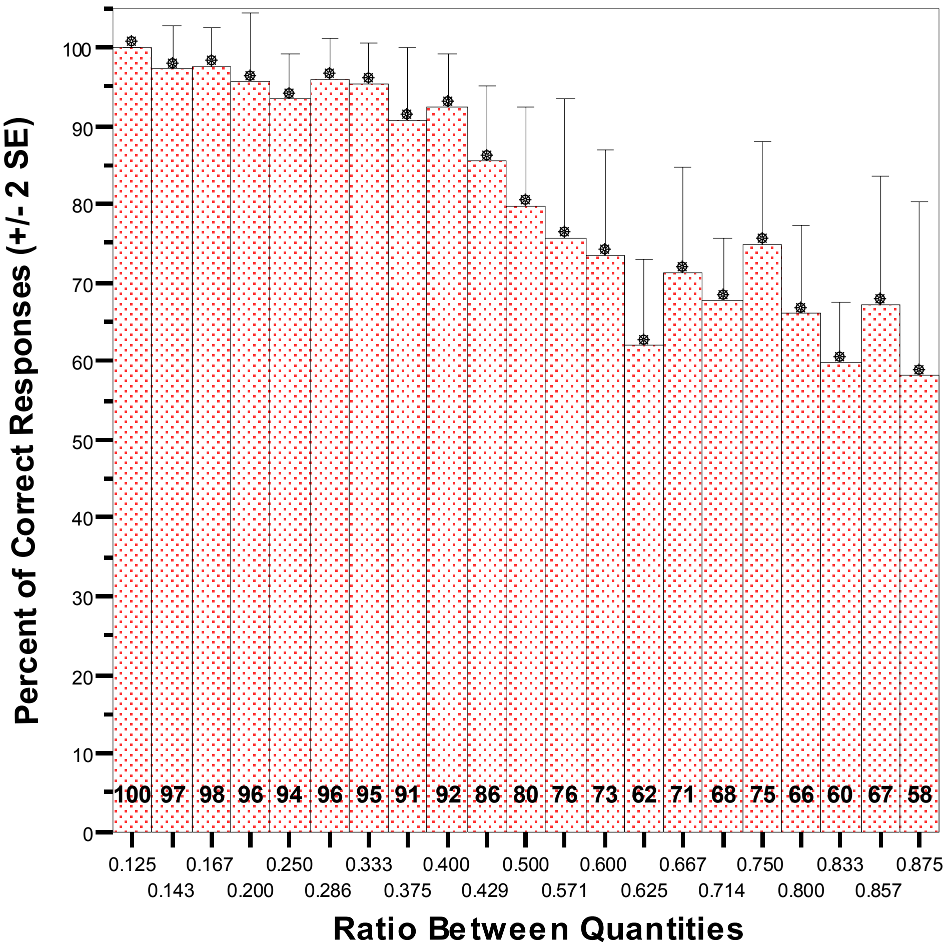


Figure 10.1. Percent of correct responses to ratios between quantities for the young.

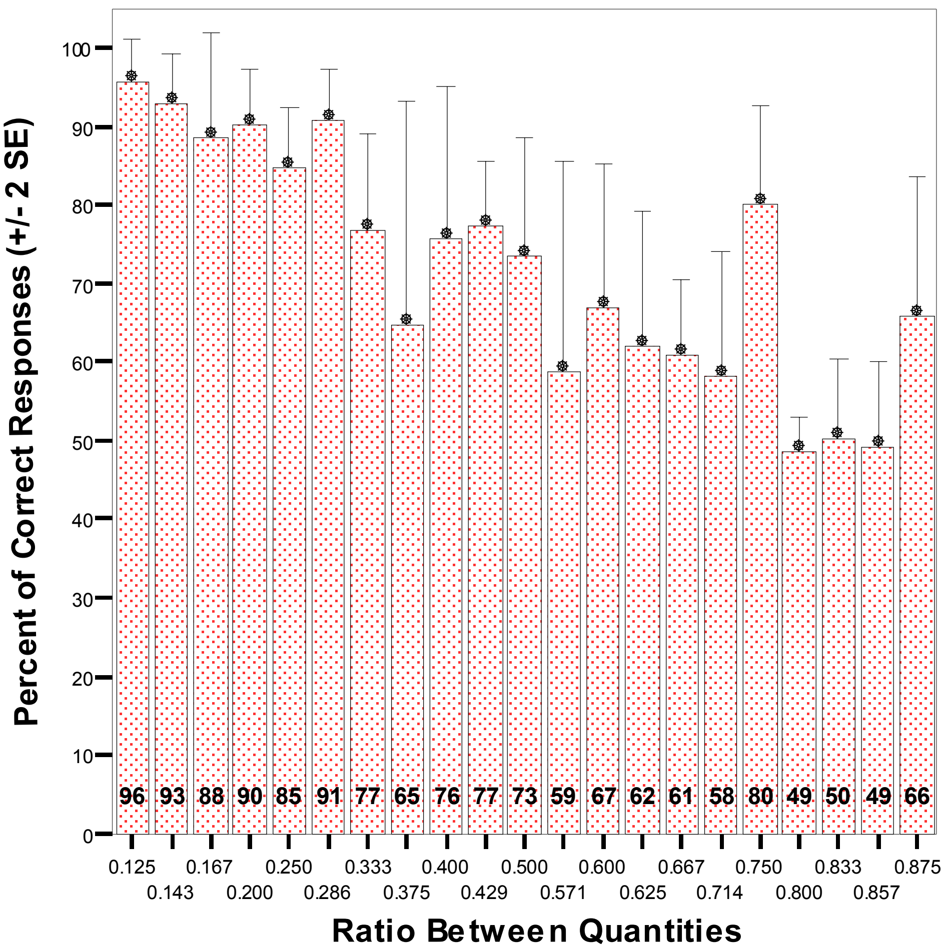


Figure 10.2. Percent of correct responses to ratios between quantities for the old.

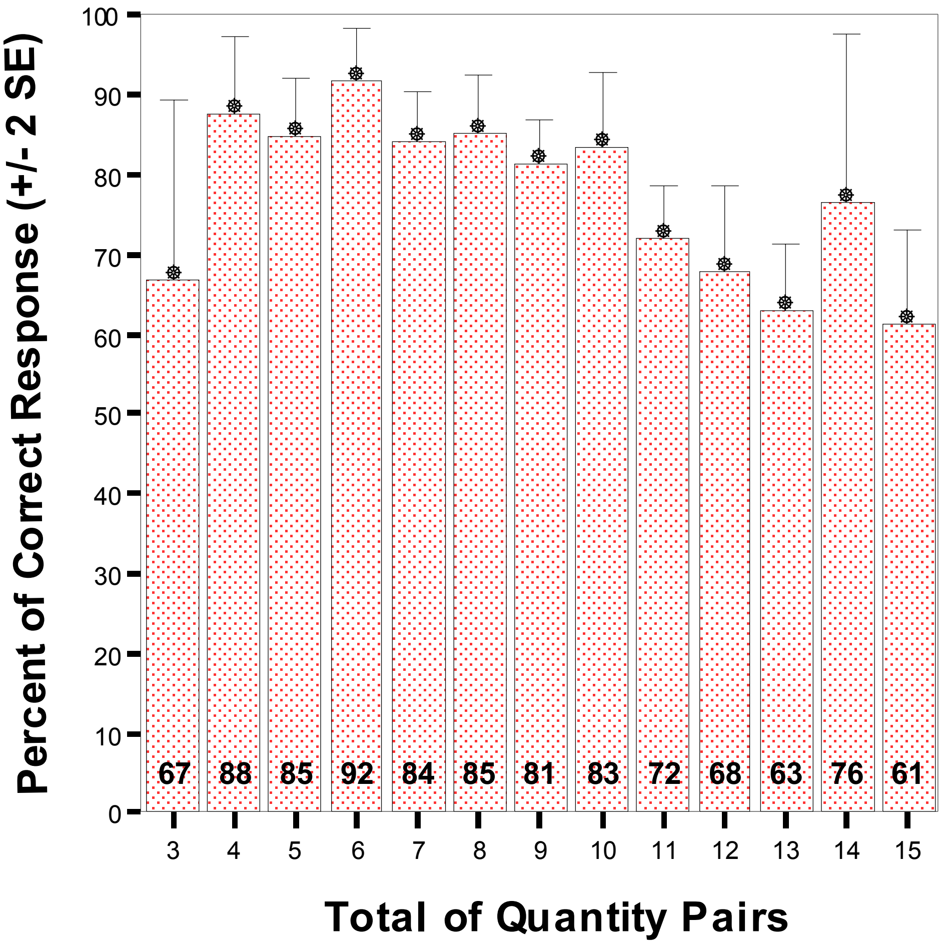


Figure 11.1. Percent of correct responses to totals of quantities for the young.

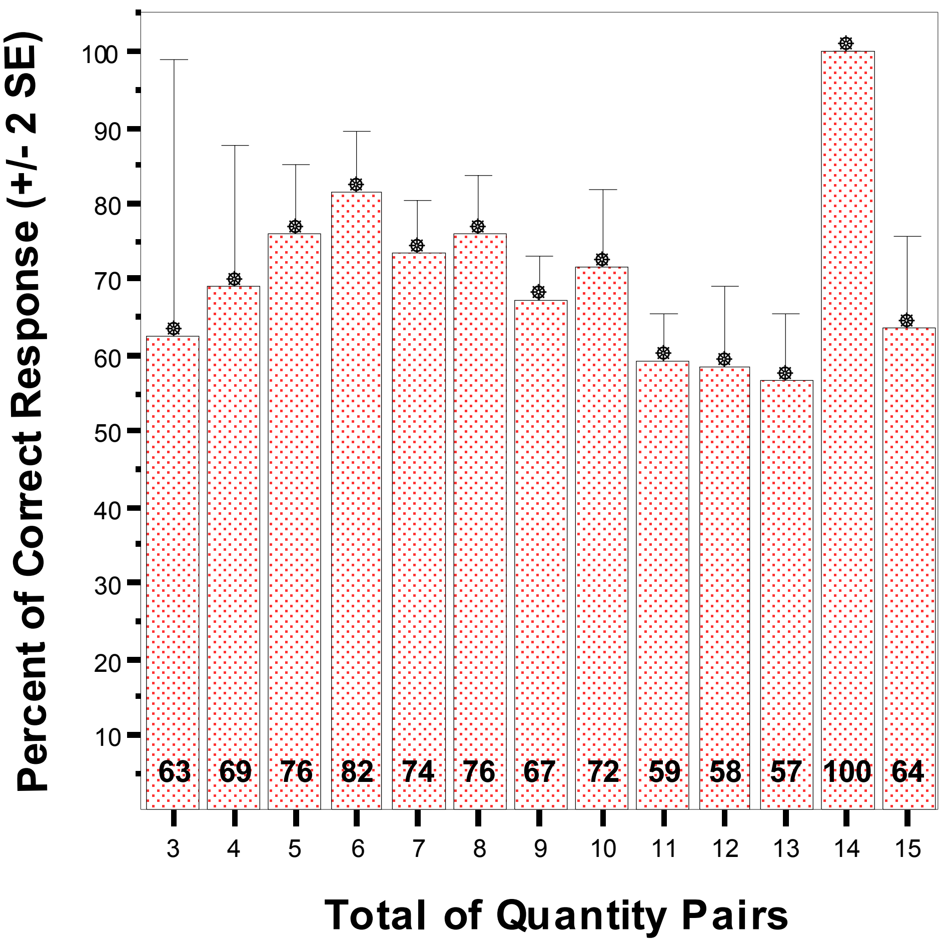


Figure 11.2. Percent of correct responses to totals of quantities for the old.

To test whether the frequency of correct responses were significantly above chance levels, binomial tests were performed for each quantity comparison and age group. A Bonferroni adjusted alpha of 0.001 was used to test significance (Table 23).

Pearson correlations between the percent of correct responses and the quantity variables were statistically significant for each age group, $p < 0.008$ in all cases. The relationship was such that the percent of correct responses increased as the difference between quantities increased for both the young (r = 0.50, p = 0.001) and the old (r = 0.43, p = 0.001). Also, the percent of correct responses increased as the ratio between quantities decreased for both the young (r = -0.60, p = 0.001) and the old (r = -0.44, p = 0.001). And the percent of correct responses increased as the total of quantities decreased for the young (r = -0.38, p = 0.001) and the old (r = -0.34, p = 0.001).

A Mann-Whitney test was performed to evaluate age differences in response time. The Mann-Whitney test was significant, N = 11, U = 1.00, p < 0.005. The mean response time for the young was significantly faster (M = 1.11, SE = 0.04) than the mean response time for the old (M = 1.97, SE = 0.06).

For both young and old age groups, response time was not correlated with any of the quantity variables, p > 0.008 in all cases. A Bonferroni adjusted alpha of 0.008 was calculated to test significance. For the young the correlation for ratio, difference, and total were r's = 0.14, 0.16, and 0.01 respectively. For the old the correlation for ratio, difference, and total were r's = 0.08, 0.09, and 0.06 respectively. Response time and the percent of correct responses was not correlated for the young (r = 0.00, p = 0.494) or for the old (r = 0.09, p = 0.251).

Table 23. Binomial tests of the frequency of correct responses to quantity comparisons by age group

				Young			Old		
Quantity Comparison	Ratio	Diff.	Total	N	Percent Correct	p	N	Percent Correct	p
1:8	**0.125**	**7**	**9**	**42**	**100.00**	**0.001**	**46**	**95.65**	**0.001**
1:7	**0.143**	**6**	**8**	**36**	**97.22**	**0.001**	**42**	**92.86**	**0.001**
1:6	**0.167**	**5**	**7**	**40**	**97.50**	**0.001**	**45**	**88.89**	**0.001**
1:5	**0.200**	**4**	**6**	**44**	**95.45**	**0.001**	**52**	**90.38**	**0.001**
1:4	**0.250**	**3**	**5**	**49**	**93.88**	**0.001**	**46**	**84.78**	**0.001**
2:7	**0.286**	**5**	**9**	**47**	**95.74**	**0.001**	**52**	**88.46**	**0.001**
2:6	0.333	4	8	8	100.00	0.004	11	81.82	0.033
1:3	**0.333**	**2**	**4**	**49**	**87.76**	**0.001**	26	69.23	0.038
3:8	**0.375**	**5**	**11**	**45**	**91.11**	**0.001**	43	72.09	0.003
2:5	**0.400**	**3**	**7**	**49**	**91.84**	**0.001**	**49**	**77.55**	**0.001**
3:7	**0.429**	**4**	**10**	**46**	**84.78**	**0.001**	**57**	**77.19**	**0.001**
4:8	0.500	4	12	5	80.00	0.188	6	66.67	0.344
3:6	0.500	3	9	20	75.00	0.021	18	83.33	0.004
2:4	**0.500**	**2**	**6**	**31**	**87.10**	**0.001**	46	71.74	0.002
1:2	0.500	1	3	18	66.67	0.119	8	62.50	0.363
4:7	**0.571**	**3**	**11**	**56**	**78.57**	**0.001**	69	68.12	0.002
3:5	**0.600**	**2**	**8**	**51**	**74.51**	**0.001**	73	65.75	0.005
5:8	0.625	3	13	66	62.12	0.032	72	62.50	0.022
4:6	0.667	2	10	14	78.57	0.029	24	58.33	0.271
2:3	**0.667**	**1**	**5**	**50**	**76.00**	**0.001**	42	66.67	0.022
5:7	0.714	2	12	73	67.12	0.002	83	57.83	0.094
6:8	0.750	2	14	17	76.47	0.025	6	100.00	0.016
3:4	0.750	1	7	49	65.31	0.022	69	60.87	0.046
4:5	**0.800**	**1**	**9**	**95**	**67.37**	**0.001**	147	48.98	0.435
5:6	0.833	1	11	100	60.00	0.028	136	50.74	0.466
6:7	0.857	1	13	72	63.89	0.012	60	50.00	0.500
7:8	0.875	1	15	70	61.43	0.036	66	63.64	0.018

Note. Bold face type indicates significant probabilities N = number of cases.
Diff. = Difference between quantities.

Mann-Whitney tests of response time by the total of quantities indicated that the young responded faster than the old for totals of 6 (U = 0.00, p = 0.002). Response times for old and young were not different for totals of 3 (U = 7.5, p = 0.089), 4 (U = 4.00, p = 0.023), 5 (U = 3.00, p = 0.015), 7 (U = 3.00, p = 0.015), 8 (U = 2.00, p = 0.009), 9 (U = 3.00, p = 0.015), 10 (U = 1.00, p = 0.005), 11 (U = 4.00, p = 0.026), and 12 (U = 3.00, p = 0.015), 13 (U = 2.00, p = 0.009), 14 (U = 3.00, p = 0.015, and 15 (U = 3.00, p = 0.015). The number of cases was eleven for each test. A Bonferroni adjusted significance level of 0.002 was used to test significance and response times are graphed in Figure 12.1.

Mann-Whitney tests indicated that response times for the young and old differed depending on certain differences between quantities. Bonferroni adjusted alpha was 0.004. Young gorillas responded faster than old gorillas for the differences of 4 (U = 0.00, p = 0.002). Response times were not statistically different for differences of 1 (U = 1.00, p < 0.005), 2 (U = 1.00, p < 0.005), 3 (U = 2.00, p < 0.009), 5 (U = 2.00, p < 0.009), 6 (U = 5.00, p = 0.041), and 7 (U = 12.00, p = 0.331). The number of cases was eleven for each test. Response times are graphed in Figure 12.2.

Mann-Whitney tests indicated that the response times for the young and old were not different for all ratios between quantities p > 0.001 in all cases (Figure 12.3).

Wilcoxon tests indicated that response time was not statistically different between correct and incorrect responses for both young (z = -0.41, p = 0.343) and old (z = -1.57, p = 0.058) gorillas. The average response time for the young for incorrect responses was 1.09 s (SE = 0.04) and for correct responses 1.17 s (SE =

0.08). For the old, the average response time for incorrect responses was 2.14 s (SE = 0.14) and for correct responses 1.89 s (SE = 0.07).

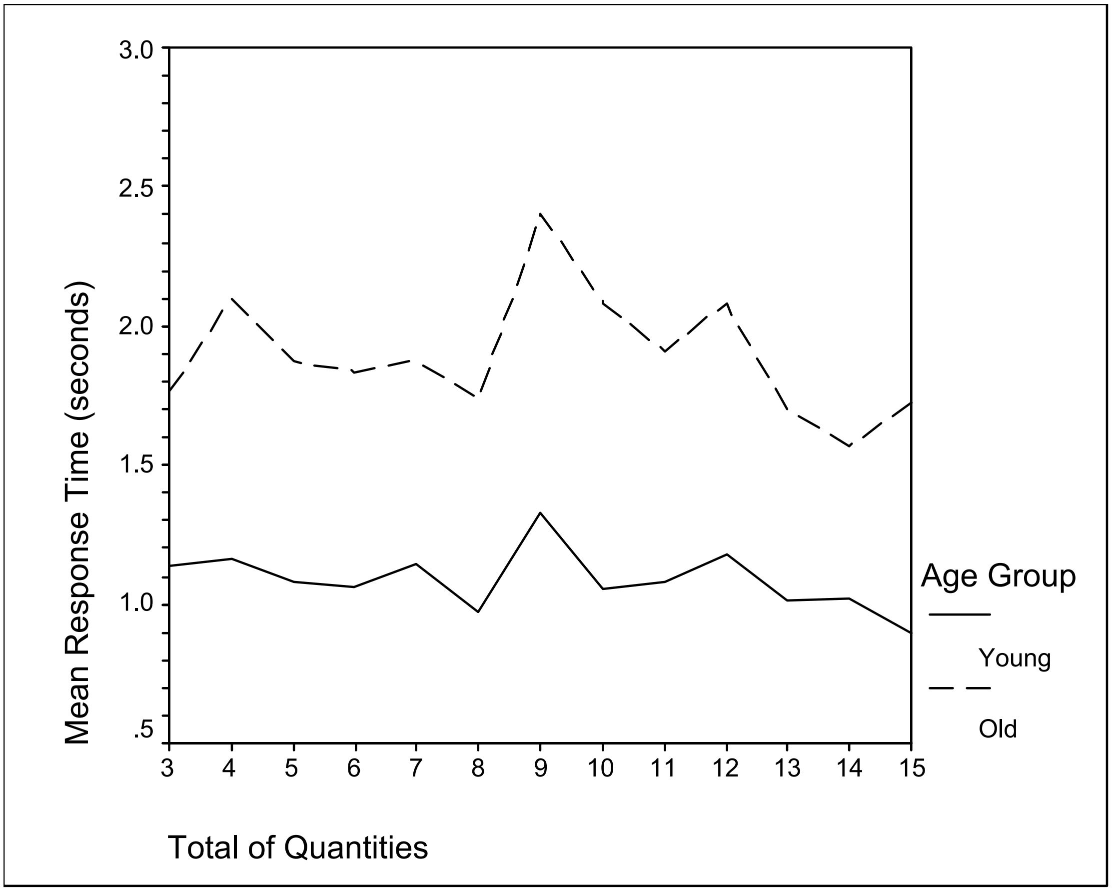


Figure 12.1. Response time to totals of quantities for the young and old.

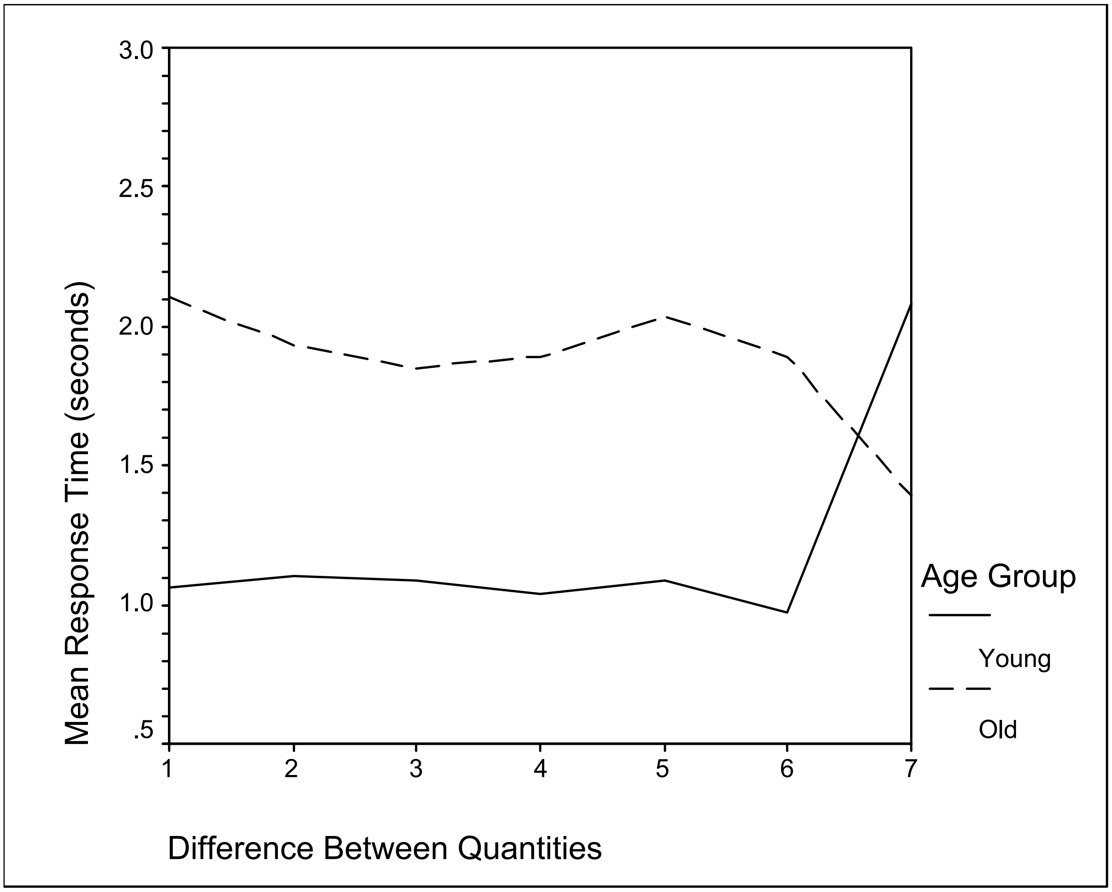


Figure 12.2. Response time to differences between quantities for the young and old.

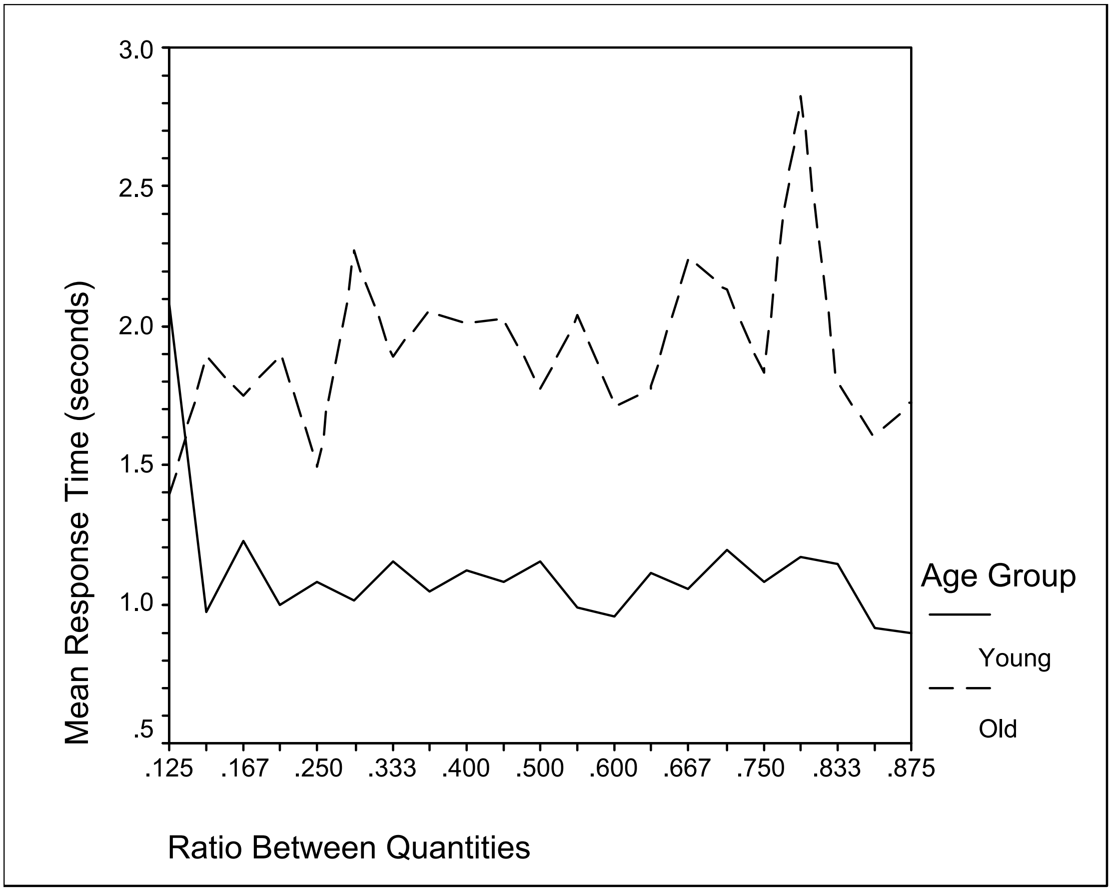


Figure 12.3. Response time to ratios between quantities for the young and old.

Wilcoxon tests indicated that response time was not statistically different between correct and incorrect responses for both young (z = -0.41, p = 0.343) and old (z = -1.57, p = 0.058) gorillas. The average response time for the young for incorrect responses was 1.09 s (SE = 0.04) and for correct responses 1.17 s (SE = 0.08). For the old, the average response time for incorrect responses was 2.14 s (SE = 0.14) and for correct responses 1.89 s (SE = 0.07).

Individual Differences

Binomial tests indicate that each individual was reliably selecting the larger quantity, p < 0.002 for each individual (Table 24).

Table 24. Binomial tests of the frequency of correct responses

Age Class	Subject	N	Percent Correct (SE)	p
Young	Charlie	243	80.66 (2.54)	0.001
	Kekla	250	75.20 (2.74)	0.001
	Kudzoo	250	78.40 (2.61)	0.001
	Stadi	250	92.86 (4.02)	0.001
	Taz	249	84.21 (8.60)	0.002
Old	Banga	250	68.40 (2.95)	0.001
	Ivan	225	69.78 (3.07)	0.001
	Katoomba	229	75.98 (2.83)	0.001
	Ozoum	242	59.92 (3.15)	0.001
	Paki	246	68.70 (2.96)	0.001
	Shamba	202	62.87 (3.41)	0.001

Note. N = number of cases.

Chi-square tests (2-sided) analyzed the frequency of correct responses by testing day for each individual. A Bonferroni adjusted alpha of 0.005 was used to test significance. The chi-square tests indicated that the frequency of correct responses differed by testing day for Banga, 2 (8, N = 250) = 27.19, p = 0.001, and Shamba, 2 (8, N = 202) = 23.40, p = 0.003 (Table 25). Individual performance by testing day is graphed in Figure 13.1 and 13.2.

Table 25. Chi-square tests of the frequency of correct responses each testing day

Age Class	Subject	df	N	Value	p (2-tailed)
Young	Charlie	9	243	10.94	0.280
	Kekla	9	250	18.360	0.031
	Kudzoo	9	250	15.68	0.074
	Stadi	9	250	12.00	0.214
	Taz	9	249	19.99	0.018
Old	Banga	9	243	27.19	0.001
	Ivan	8	225	9.91	0.272
	Katoomba	9	229	10.49	0.312
	Ozoum	9	242	10.54	0.308
	Paki	9	246	16.60	0.055
	Shamba	8	202	23.40	0.003

Note. N = number of cases.

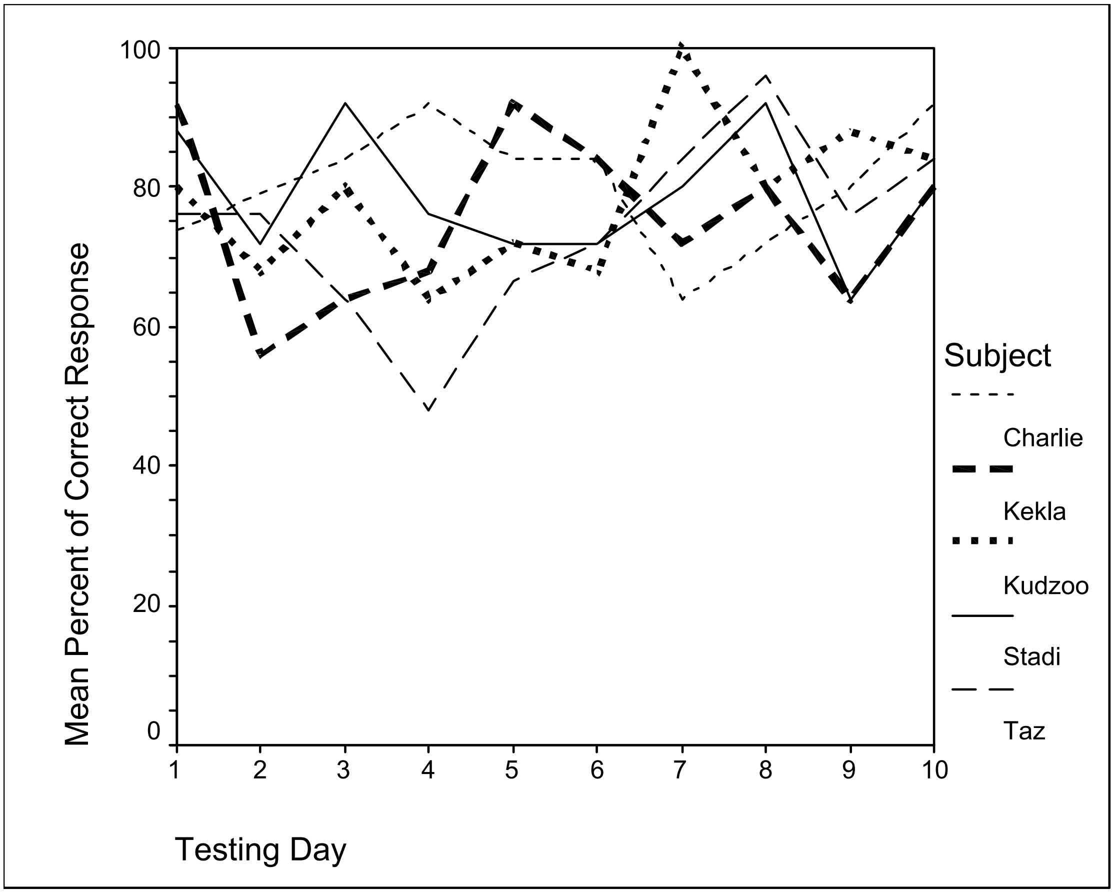


Figure 13.1. Percent of correct responses for the young each testing day.

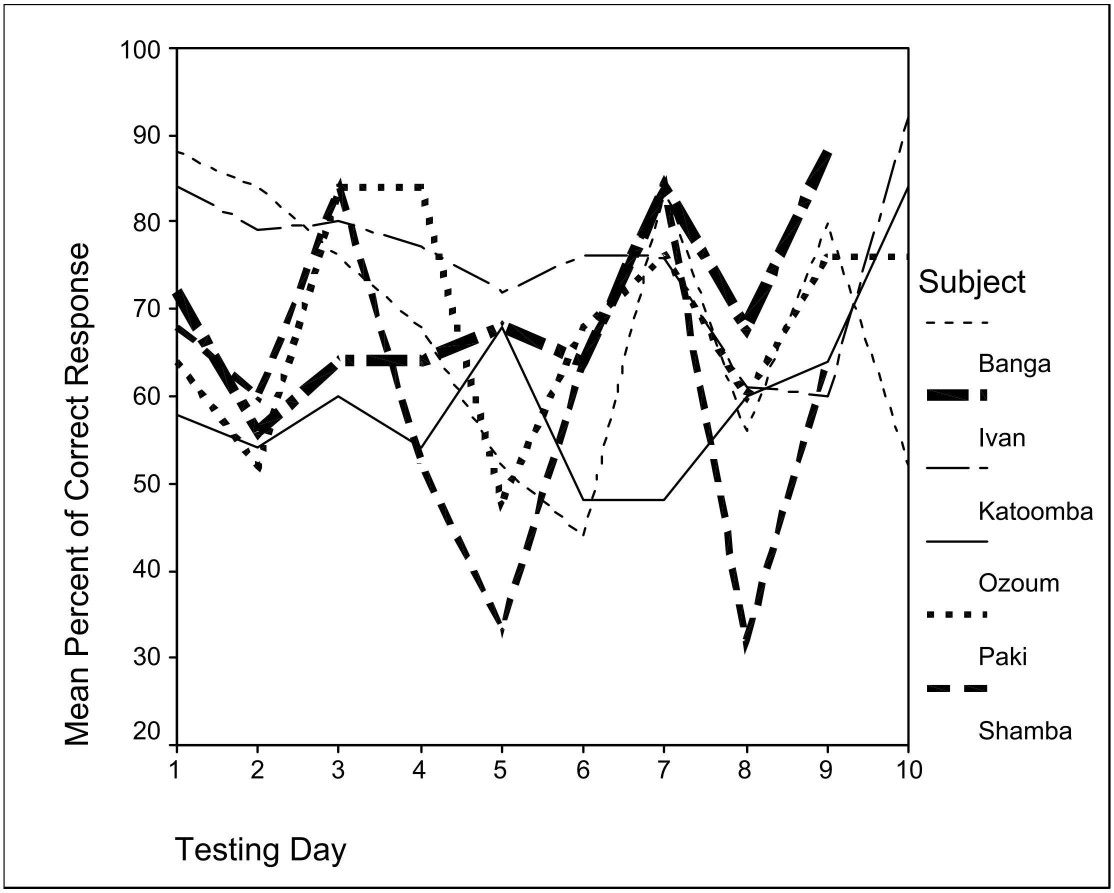


Figure 13.2. Percent of correct responses for the old each testing day.

Pearson correlations between the percent of correct responses and the quantity variables indicated significant correlations for certain individuals. A Bonferroni adjusted alpha of 0.008 was used for significance. The relationship was such that percent of correct responses increased as the ratio between quantities decreased for Charlie, Kekla, Kudzoo, Stadi, Taz, Banga, Ivan, and Katoomba, $p < 0.008$ in all cases. The percent of correct responses increased as the difference between quantities increased for Charlie, Kekla, Stadi, Taz, Banga, Ivan, Katoomba, and Paki, $p < 0.008$ in all cases. And the percent of correct responses increased as the total of quantities decreased for Stadi, Banga, and Katoomba, $p < 0.008$ in all cases (Table 26).

Table 26. Pearson correlations between percent of correct responses and the quantity variables

Age Group	Subject	Ratio	Difference	Total
Young	Charlie	-0.56*	0.53*	-0.44
	Kekla	-0.61*	0.49*	-0.41
	Kudzoo	-0.50*	0.27	-0.23
	Stadi	-0.73*	0.75*	-0.51*
	Taz	-0.66*	0.58*	-0.40
Old	Banga	-0.59*	0.48*	-0.51*
	Ivan	-0.70*	0.59*	-0.32
	Katoomba	-0.49*	0.53*	-0.57*
	Ozoum	-0.37	0.30	-0.06
	Paki	-0.30	0.49*	-0.35
	Shamba	-0.40	0.43	-0.45

Pearson correlations for response time and the quantity variables indicated significant correlations for certain individuals. An adjusted alpha of 0.008 was used to test significance. The relationship was such that response time increased as the difference between quantities increased for Stadi and Paki, $p < 0.008$ in both cases. Also response time increased as the difference between quantities decreased for Taz and Katoomba, $p < 0.008$ in both cases. Response time increased as the ratio between and the total of quantities decreased for Paki, $p < 0.008$ (Table 27). Response time was not correlated for any individual with the percent of correct responses, $p > 0.01$.

Table 27. Pearson correlations between response time and the quantity variables

Age Group	Subject	Ratio	Difference	Total	Percent Correct
Young	1	-0.07	0.06	-0.22	-0.25
	2	-0.30	0.47**	0.07	0.15
	3	0.11	-0.48**	0.26	-0.24
	4	-0.26	-0.075	-0.02	0.06
	5	0.07	-0.02	-0.04	0.09
Old	6	0.25	-0.08**	0.51	-0.29
	7	0.40	-0.29	0.24	-0.30
	8	-0.49**	0.53**	-0.57**	0.00
	9	0.15	-0.14	-0.38	-0.10
	10	-0.30	0.00	-0.07	0.21
	11	0.06	0.03	0.20	-0.23

*$p < 0.01$.
**$p < 0.008$.

CHAPTER IV

DISCUSSION

Relative Numerousness Judgments Without Prior Specific Training

Contrary to the hypothesis that the gorillas would reliably make relative numerousness judgments without prior specific training to do so (experiment 1), the findings indicate that only four (2 young and 2 old) of the eleven gorillas reliably performed relative numerousness judgments before specific training to do so. The percent of correct responses for these four gorillas averaged at 69%, 70%, 76%, and 77% correct. The seven gorillas (3 young and 4 old) that did not select the larger quantity more often than chance averaged from 49% to 62% correct. Nine of the eleven gorillas, including one old and one young that were reliably performing relative numerousness judgments, performed more poorly than chimpanzees and orangutans tested on identical or very similar relative numerousness judgment tasks without specific prior training to do so.

Rumbaugh and colleagues (1987) found that two chimpanzees (Sherman and Austin) performed relative numerousness judgments from the first day of testing when given the same quantity comparisons used in the present study. Their performance averaged at 97 % and 96% correct. Dooley and Gill (1977) found that one chimpanzee (Lana) performed relative numerousness judgments to reliably select the larger quantity from the smaller quantity with her performance averaging 94% correct for the same quantity comparisons used in the current study and additional comparisons of 1:5, 2:5, 3:5, and 4:5. Pérusse and Rumbaugh (1990) presented two chimpanzees (Sherman and Austin) with quantity comparisons of 5:6, 5:7, 5:8, 6:7, 6:9, 7:8, 7:9, 8:9, and 9:10 and found that their performance averaged 90% and 79% correct. Beran (2001) presented two chimpanzees (Sherman and Lana) with the same quantity comparisons used in the current study and additional comparisons of 1:5, 2:5, 2:6, 3:5, 3:6, 3:7, 4:5, 4:6, 4:7, 4:8, 5:6, 5:7, 5:8, 5:9, 6:7, 6:8, 6:9, 7:8, 7:9, and 8:9 and found that both of the two chimpanzees were able to perform relative numerousness judgments with their performance averaging 82% and 73% correct. Call (2000) presented three orangutans (Chantek, Teriang, and Solok) with the same quantity comparisons used in the current study and additional quantity comparisons of 1:5, 2:5, 2:6, 3:5, 3:6, 4:5, 4:6, and 5:6 and found that all three orangutans performed relative numerousness judgments with their performance averaging 95%, 86%, and 82% correct.

One possible explanation of the gorillas' poorer performance in comparison to the previously established performance of three chimpanzees and three orangutans is of experimental history. Before the initial relative numerousness tests of Lana by Dooley and Gill (1977) and of Sherman and Austin by Rumbaugh and colleagues (1987), these three chimpanzee subjects were extensively involved in language acquisition projects. Dooley and Gill (1977) write that Lana's language type skills of naming objects and people, labeling same versus different, and describing prepositional relationships may have provided her with prerequisites to learning relative number concepts. And Rumbaugh and colleagues (1987) write that Sherman and Austin may have learned a preference for the larger quantity from a previous experiment. The orangutans Chantek, Teriang, and Solok also possessed prior extensive experimental histories in cognition and learning tasks and Chantek was a participant in a language acquisition project. The gorillas in the current study have participated in operant conditioning, and some learning and memory tasks, however not to the same extent as the three previously mentioned chimpanzees and orangutans. Thus their poorer performance may be related to their more limited experimental histories.

Two young and two old gorillas reliably made relative numerousness judgments for the 1:4 comparisons, performance averaged 88% to 100% correct, and one old animal for the 1:3 comparisons, performance averaged 92% correct. Both 1:4 and 1:3 are represented by smaller ratios and larger differences in the set of quantity comparisons. In general, the group's performance was better for quantity comparisons composed of smaller ratios and larger differences between quantities. Smaller ratios and larger differences between quantities indicate high disproportion or dissimilarity between the two quantities. The total of quantities was not linearly related to percent of correct responses. That the ability to discriminate between two numerosities or quantities improves as the numerical distance (the distance between quantities) increases is referred to as the numerical distance effect (Dehaene et al., 1998).

Dooley and Gill (1977) report that Lana the chimpanzee's performance averaged 100% correct for 1:3 and 98% for 1:4 and the authors noted that most of her errors occurred at large ratios and small differences for their one chimpanzee subject. Call (2000) reported the average performance of the three orangutans at 100%, 92%, and 92 % for 1:3 comparisons and 100% for 1:4 and the author also reported that all three individual orangutans expressed significant negative correlations with the ratio between quantities and significant positive correlations with the difference between quantities. Pérusse and Rumbaugh (1990) noted that performance was better for their two chimpanzees for smaller ratios between quantities. By comparing the percent of correct responses to 1:4 and 1:3 for the four gorillas reliably performing relative numerousness judgments to the chimpanzees and orangutans, one notices that their average performance was not much different.

When the individual correlations of the gorillas were examined, the percent of correct responses was correlated with ratio and difference for only three (two young and one old) of the eleven gorillas and the total of quantities was not correlated with any individual. The lack of significant correlations in seven of the gorillas indicated that there is some variability in the relation of percent of correct responses to ratio, difference, and the total of/between quantities. Similarly, individual differences were found by Beran (2001) in the correlations between percent of correct responses and the quantity variables. Both of the two chimpanzees expressed significant negative correlations to ratio and total, however only one expressed a significant positive correlation to the difference between quantities. While the ratio and the difference between quantities was not correlated for seven gorillas, the direction of the effect was consistent in the three gorillas expressing significant correlations; better performance for smaller ratios and better performance on larger differences.

Also contrary to the hypothesis, an age-related difference was not found for relative numerousness judgments without prior specific training to do so. The findings indicated that young and old gorillas chose the larger quantity at the same frequencies, 65% and 61% respectively. Differences were not found between the young and old in how often they selected the larger quantity depending on the ratio, difference, and total of the quantity comparisons. This finding was not unexpected because the majority of gorillas responded at chance levels when selecting a quantity and some previous research of learning and memory in apes suggests that advancing age in apes is not associated with global deficits, thus the finding of similar performance between young and old gorillas was not atypical.

Response time also did not differ between young and old gorillas during the test of relative numerousness judgments without prior training to do so. Young and old gorillas responded at similar speeds, 2.07 seconds for the young and 2.85 for the old. Other authors report that old and young monkeys did not differ in color discrimination and reversal learning (Cohen, Eisdorfer, & Bowden, 1979), delayed-response tasks (Davis, 1978), or in response time tasks (Baxter & Voytko, 1996). In addition, correct responses required the same amount of time as incorrect responses for both the young, old, and the individual. Even though some gorillas responded faster or slower than other gorillas, response time was not linearly related to the ratio, difference, or total between/of the quantities, or the percent of correct responses for the young and the old or for any individual. These response time findings of no difference between young and old gorillas are not unexpected in light of the findings that individually only four gorillas were reliably performing relative numerousness judgments. The majority of gorillas were not performing relative numerousness judgments, thus age-differences in response times may be absent or undetectable.

Relative Numerousness Judgments With Specific Training

After specific training using a differential reinforcement contingency in which subjects were not reinforced for selecting the smaller quantity, experienced an ITI of 30 seconds for incorrect responses, and had to reach a criterion performance level of 80% for two consecutive testing days (experiment 2), each individual gorilla reliably performed relative numerousness judgments to select the larger quantity from the smaller quantity. Individuals were reliably performing relative numerousness judgments when their performance was considered across all of the testing days they required to reach the criterion; the percent correct for each individual averaged from 66% to 93% (only one gorillas was under 75%).

When considering only the last two testing days at criterion levels, the percent correct ranged from 82% to 98%. The performance of the gorillas for the last two days after training can be considered near-asymptotic levels of performance, as the average percent of correct responses were well within the average performance ranges reported by previous experimenters for chimpanzees and orangutans. The average percent of correct responses at near-asymptotic levels were a marked increase from the performance exhibited without training in experiment 1. It is possible that the training procedure alone was not accountable for the establishment of reliable relative numerousness judgments in each gorilla. Instead, reliable relative numerousness judgments may have resulted from simply receiving additional trials and not from the contingencies of experiment 2. However, this alternate explanation for improved performance is not likely.

After specific training, the group as a whole reliably performed relative numerousness judgments to each of the quantity comparisons. Individually, seven gorillas (three young and four old) were reliably performing relative numerousness judgments to certain quantity comparisons. All seven gorillas selected the larger quantity for quantity comparisons of 1:4, two of the three young gorillas to 1:3, one of the three young gorillas to 2:4, one of the four old and one of the three young gorillas to 2:3, and one of the four old gorillas to 3:4. For these seven gorillas performance averaged from 76% to 100% correct. Four gorillas (two young and two old) did not respond correctly more often than chance for any quantity comparison. However, this finding was more likely the result of insufficient trials to obtain significance because these four gorillas reached the criterion in two or three testing days. Also, four of these three gorillas obtained upwards of 80% correct, and in many cases 100%, for the majority of the quantity comparisons.

At near-asymptotic performance as a group, the percent of correct responses increased as the difference between quantities increased. The ratios between quantities and the totals of quantities were not linearly related to the percent of correct responses. When examining the individual correlations of subjects for performance during the last two testing days, the percent of correct responses was not correlated with ratio, difference, or total for any individual. Again, this may indicate individual variability within the relation between reliable numerousness judgments and the ratio, difference, and the total of/between quantities. Interestingly, in experiment 1 during relative numerousness judgments without prior specific training, as a group both ratio and difference were linearly correlated with percent of correct responses.

There was not an age difference in the ability to perform relative numerousness judgments when considering the performance of the young and old for the first day of testing (75% and 74% respectively) or for the last two days of testing at near-asymptotic performance levels (88% and 87% respectively). Also, young and old gorillas required the same number of days and the same number of trials to reach the criterion. The young averaged 4 days and 58 trials and the old averaged 5 days and 82 trials. As in experiment 1, in experiment 2 at near-asymptotic performance levels differences did not exist between the young and old in how often they selected the larger quantity in regard to the ratio, difference, and total of/between the quantity comparison.

Young and old gorillas responded at the same speed the first day of testing, 1.96 and 2.60 seconds respectively, and at near-asymptotic performance levels during the last two days of testing, 1.28 and 2.36 seconds respectively. At near-asymptotic performance levels, response times were different between individual gorillas with some responding faster or slower than others for the last two testing days, but response times for correct and incorrect responses did not differ for the young, old, within individuals, or for the group as a whole. Moreover, response time for the last two testing days was not correlated with any of the quantity variables for the young, old, the individual, or for the group as a whole.

Summation Without Specific Prior Training

In regard to summation, the findings indicated that each individual gorilla was reliably performing summation operations to select the larger paired quantity from the smaller without specific prior training to do so (experiment 3). The percent of correct responses averaged from 93% to 75% for the young and 60% to 76% for old gorillas. The percent of correct responses obtained in the current summation experiment are within the ranges reported in chimpanzees by other authors. Summation performance reported in two chimpanzees on the first two of four summation tests averaged approximately 60% to 87% correct (exact averages were not reported) between chimpanzees for quantity comparisons similar to the comparisons used in the present study (Rumbaugh et al., 1987). On the last summation test with the quantity of five added to the existing comparisons, the percent of responses averaged approximately 87% to 95% correct (exact averages were not reported) between the same two chimpanzees (Rumbaugh et al., 1987). In subsequent summation tests in the same two chimpanzees previously discussed, the percent of correct responses averaged 92% and 95% (Rumbaugh et al., 1988) and 82% and 72% correct (Pérusse & Rumbaugh, 1990). Beran (2001) reported that two chimpanzees averaged 80% and 74% correct in summation tests for similar quantity comparisons.

Accordingly with the hypothesis of age-differences existing for summation ability, the findings indicated that the young selected the larger quantity more often than the old did. Tests of summation ability indicated that the young (77% correct) performed better than the old (68% correct) gorillas. The pattern of performance indicated that both young and old were reliably performing summation operations to select the larger quantity for quantity comparisons of 1:4, 1:5, 1:6, 1:7, 1:8, 2:5, 2:7, and 3:7. In addition, the young were also reliably discriminating quantity comparisons of 1:3, 2:3, 2:4, 3:5, 3:8, 4:5, and 4:7. The pattern of performance was such that young and old gorillas selected the larger quantity pair just as often to different ratios between quantities. However, when the quantity comparisons were defined by the differences between quantities, the young selected the larger paired quantity more often than did the old for differences of 1, 2, and 5. When defined by the total of quantities, the young selected the larger pair quantity more often than the old for quantity totals of 9. By examining the average percent correct for old and young for the different quantity comparisons, one finds that the old averaged better than the young only for four of the twenty-seven quantity comparisons. Thus, in general the young performed better than the old and also for some specific differences and totals. The overall age-difference in summation performance was because of this general tendency of the young to perform better and because the young performed better for certain differences and totals of/between quantities.

Correlations of the percent of correct responses and ratio, difference, and total of/between quantities revealed that for both the old and the young the percent of correct responses increased with increasing differences and decreasing ratios and totals during summation. The strength of the relationship, indexed by the Pearson correlations, with ratio, difference, and total was greater for the young than the old for all three cases of the quantity variables, thus the old are least affect by the quantity variables. The individual correlations indicated that the percent of correct responses increased as the ratio decreased for eight gorillas (five young and three old). The percent of correct responses increased as the difference increased for eight gorillas also (4 young and 4 old). And the percent of correct responses increased as the total decreased for three gorillas (1 young and 3 old). And for two old gorillas, the ratio, difference, or total of/between quantities were not related to percent of correct responses. These correlational results are similar to those found by Beran (2001) in two chimpanzees. The author found that for both chimpanzees that the percent of correct responses increased with smaller ratios and larger differences. And for one chimpanzee the percent of correct responses increased for smaller totals. Rumbaugh and colleagues (1987, 1988) also described a trend of better summation performance for smaller ratios between quantities.

It may be that smaller ratios, larger differences, and smaller totals of/between quantities make discrimination easier. We found that the young gorillas were more affected by the quantity variables therefore the finding of their overall better performance may be related to the stronger affect of the three quantity variables. Another possible explanation of these trends in both the young and the old is motivation. When the paired quantities differed by large amounts (small ratios and large differences), for example 1:8, the difference in the amount of reward for selecting the larger quantity pair may have prompted stronger preferences to discriminate. Quantities that differed very little (large ratios and small differences), for example 1:2 may not have prompted a strong preference in animals to discriminate. Also, when the paired quantities both represented large totals, for example 6:7, a weak preference to discriminate may have existed. To account for age differences, the young gorillas may have possessed overall stronger preferences to discriminate between quantities than did the old gorillas.

In contrast with experiments 1 and 2, in the test of summation ability significant correlations existed for the majority of individuals. This may have resulted because of the increase in quantity comparisons from ten comparisons in experiment 1 and 2 to twenty-seven comparisons in experiment 3. The increase in quantity comparisons also increased the range of ratios, differences, and totals of/between quantities which may have allowed for a greater effect of the quantity variables on discrimination ability. Also, this difference may reflect a discontinuity in the processes of relative numerousness judgments and summation such that the ratios, differences, and totals of/between quantities direct a greater effect on the processes of summation.

Age differences also existed in response times with the young responding faster at 1.11 seconds than the old gorillas at 1.97 seconds. These results may be comparable to the finding in humans of older adults exhibiting slower response speeds during simple addition and subtraction tasks (Geary et al., 1993; Salthouse & Kersten, 1993; Silwinski et al., 1994; Geary & Lin, 1998) and for older monkeys responding more slowly in a response time task (Bachevalier et al., 1991). The pattern was such that the young responded faster than old gorillas to totals of 6 and differences of 4. Individual correlations reveal that one young and one old gorilla exhibited slower response times for larger differences between quantities. The opposite pattern was observed in one young and one old gorilla, slower response times were associated with smaller differences. And for one old gorilla slower response times were associated with smaller ratios and totals of/between quantities. The implications of these individual correlations are unclear however, response time was not linearly correlated with ratio, difference or the total of/between quantities for age groups.

Relative Numerousness Judgments, Summation, and Subitizing

The results obtained from the gorillas for both relative numerousness judgments with and without specific prior training did not strongly support the presence of subitization. First, accordingly with the human subitizing literature fast responding should be associated with better accuracy as subitizing is described as fast and accurate and estimation as less fast and less accurate than subitizing. For both experiments 1 and 2, fast responding was not associated with better accuracy for the individual, group, or for the young or old. For one old gorilla, response times were actually faster as performance declined. Nor did response time differ between correct and incorrect responses for the individual, group, or the young and old.

Second, visually examining response times for trends did not indicate patterns paralleling the performance reported for numerousness tasks in humans, monkeys, or apes. Specifically, visual examination indicated near uniform response times across the range of quantity totals (three to seven) for both young and old gorillas in relative numerousness judgments without training to do so. And at asymptotic performance levels for relative numerousness judgments with specific training, visual examination indicated near uniform response times across the range of quantity totals for the young, but for the old a large decrease in response time from three to four quantity totals followed by nearly uniform response times across the remaining set of quantity totals (four to seven). This finding conflicts with the previously found speeds of responding to numerousness tasks in humans and one chimpanzee by Kaufman et al. (1949), Mandler and Shebo (1982), Murofushi (1997), and Tomonaga and Matsuzawa (2002).

Performance in humans for limited exposure duration of stimuli was characterized by shallow or zero response time slopes to numerosities from one to three or four and increased and constant response slopes to numerosities from six to the maximum array numerosity (Kaufman et al., 1949; Mandler & Shebo, 1982; Tomonaga & Matsuzawa, 2002). During limited exposure to the sample and comparison stimuli one chimpanzee's response time was characterized also by shallow or zero response time slopes to numerosities from one to four and increased response time slopes to numerosities of four to six, and then from six to nine response time remained constant at the increased level (Tomonaga & Matsuzawa, 2002). In humans exposed to unlimited duration of stimuli was characterized also by shallow or zero response time slopes to numerosities from one to three (Mandler & Shebo, 1982) or four (Tomonaga & Matsuzawa, 2002) and response time slopes from four to the maximum array numerosity were a straight increasing line. The performance of one chimpanzee for unlimited exposure to the stimuli was characterized also by shallow and zero response time slopes for numerosities from one to three (Murofushi, 1997) or five (Tomonaga & Matsuzawa,

2002) followed by increased response time up to the second largest numerosity in the set and then decreased response time to the largest numerosity in the set.

The gorillas' performance most closely resembled the performance found in humans by Thomas, Philips, and Young (1999). Thomas and colleagues (1999) found uniform response times over the range of numerousness arrays (one to eleven) in humans discriminating dots. Our results may match Thomas and colleagues (1999) because of methodological similarities, specifically Thomas and colleagues (1999) presented subjects with a limited number of arrays, whereas Kaufman and colleagues (1949) and Mandler and Shebo (1982) used more than fifteen different numerosity arrays. A most relevant methodological difference was that this study required simultaneous discriminations of two quantities for unlimited exposure durations of stimuli. The typical numerousness and subitizing procedure is of consecutive presentation of numerosity arrays (Kaufman et al., 1949; Mandler & Shebo, 1982; Murofushi, 1997; Thomas et al., 1999; Tomonaga & Matsuzawa, 2002) and for only limited exposure durations of the numerosity arrays (Kaufman et al., 1949; Thomas et al., 1999).

The results obtained from the gorillas during the test of summation ability did not strongly support the presence of subitization. Visual examination of the trend between response speed and the total of quantities during summation indicated uniform speed of responding for the young and nearly uniform response speed in the old with an increase in response time for totals of eight to nine quantities. This pattern of response speed does not resemble the pattern indicating subitization in one chimpanzee and in humans (Kaufman et al., 1949; Murofushi, 1997; Mandler & Shebo, 1982; Tomonaga & Matsuzawa, 2000). Rather the pattern in gorillas was again more similar to the results found in humans by Thomas and colleagues (1999) of uniform response times across increasing numerosities.

Conclusions

Based on past research on the numerical competencies of nonhuman primates for relative numerousness judgment, I expected the results to reveal the gorillas to reliably make relative numerousness judgments to choose the greater quantity before specific training to do so. However, the gorillas did not choose the larger quantity until after specific training to do so. Undoubtedly, even in captivity the gorillas had encountered opportunities to perform relative numerousness judgments before testing, for example, personal observations reveal the gorillas prefer to visit the feeding site with the most food. Yet, even with their natural relative numerousness judgment experience, the majority of gorillas did not perform relative numerousness judgments before receiving specific training to do so. Interestingly, individual gorillas began performing relative numerousness judgments rather quickly after corrective methods and a differential reinforcement contingency were implemented. This suggests not that the gorillas *could* not perform relative numerousness judgments before receiving specific training, but that they *did* not performing relative numerousness judgments until after the contingencies were changed.

I speculate that for the gorillas the consequences of choosing the smaller quantity was not sufficiently strong enough to support choosing the larger quantity. Choosing the smaller quantity during experiment 1 was immediately followed by an opportunity to receive more food in a new trial and the smaller and larger quantity differed by only a maximum of three food items. This may represent a species difference in performance between the gorillas of the current study and the chimpanzees and orangutans previously discussed as the chimpanzees and orangutans reliably performed relative numerousness judgments without prior specific training to do so or be related more so to experimental history.

Based on some reports in humans and apes from the aging literature, the old gorillas were predicted to exhibit poorer and slower performance when compared to the young gorillas. Instead, the old gorillas performed more poorly and more slowly than did the young gorillas only during summation tasks. The pattern of age differences during summation tasks was such overall old gorillas performed more poorly and more slowly than did their younger counterparts and also for specific ratios, differences, and totals of/between quantities during summation tasks. Age differences may have been found for summation tasks for several reasons. Firstly, summation may be a more complex skill or involve more elaborate processes than relative numerousness judgment. Thus, because of increased task elaboration or complexity, the old may have performed more poorly and more slowly than the young gorillas. Whether summation was a more elaborate or complex task for the gorillas to perform than relative numerousness tasks was not apparent from this data.

Secondly, relative numerousness judgments may have remained invariant and summation ability may have deteriorated across advancing age in the gorillas because relative numerousness judgments might be a more fundamental ability in the lives of animals. Relative numerousness judgments are described as fundamental in the daily activity of foraging. It is not hard to imagine the importance of discriminating the food resource or patch with the most food from those with lesser amounts; discriminating would maximize the benefits associates with feeding. In addition to foraging efficiency, the ability to judge the relative numerousness of an opposing group of contra- or con-specifics before initiating potential competitive acts may be adaptively significant (Hauser, 1997). Thus, the adaptive significance of relative numerousness judgments may facilitated the retention of this ability across advancing age. Summation might also be involved in foraging such that when an animal encounters two or more patches of food each of which is composed of subpatches of food, selection of the largest patch would require the combining of subpatches (Olthof et al., 1997). However, summation is probably not as instrumental in foraging efficiency as relative numerousness judgments as the frequency of encountering food patches needing subpatch combination to determine the larger food patch is probably lower than simply encountering food patches not composed of subpatches. The major implication of the age-differences found in summation ability in gorillas indicates that subject age should be considered as a variate in number-related and other cognition tests in gorillas and other great apes.

REFERENCES

Albert, M.S., & Moss, M.B. (1996). Neuropsychology of aging: Findings in humans and monkeys. In E.L. Schneider, and J.W. Rowe (Eds.), *Handbook of the Biology of Aging* (4th ed., pp. 217-233). New York, NY: Academic Press.

Anderson, J.R., Awazu, S., & Fujita, K. (2000). Can squirrel monkeys (*Saimiri sciureus*) learn self control? A study using food array selection tests and reverse-reward contingency. *Journal of Experimental Psychology, 26,* 87-97.

Bachevalier, J. (1993). Behavioral changes in aged rhesus monkeys. *Neurobiology of Aging, 14,* 619-621.

Bachevalier, J., Landis, L.S., Walker, L.C., Brickson, M., Mishkin, M., Price, D.L., & Cork, L.C. (1991). Aged monkeys exhibit behavioral deficits indicative of widespread cerebral dysfunction. *Neurobiology of Aging, 12*, 99-111.

Bakner, L, & Treichler, F.R. (1993). Age differences in the acquisition and retention of concurrent discriminations by monkeys. *Animal Learning & Behavior, 21(1),* 51-58.

Bartus, R.T., Dean, R.L., III, & Beer, B. (1980). Memory deficits in aged Cebus monkeys and facilitation with central cholinomimetics. *Neurobiology of Aging, 1,* 145-152.

Bartus, R.T., Dean, R.L., III, & Fleming, D.L. (1979). Aging in the rhesus monkey: Effects on visual discrimination learning and reversal learning. *Journal of Gerontology, 34,* 209-219.

Bartus, R.T., Fleming, D.L, & Johnson, H.R. (1978). Aging in the rhesus monkey: Debilitating effects on short-term memory. *Journal of Gerontology, 33,* 858-871.

Baxter, M.G., & Voytko, M.L. (1996). Spatial orientating of attention in adult and aged rhesus monkeys. *Behavioral Neuroscience, 110(5),* 898-904.

Beran, M.J. (2001). Summation and numerousness judgments of sequentially presented sets of items by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 115(2),* 181-191.

Beran, M.J., Pate, J.L., Richardson, W.K., & Rumbaugh, D.M. (2000). A chimpanzee's (*Pan troglodytes*) long-term retention of lexigrams. *Animal Learning & Behavior, 28(2),* 201-207.

Beran, M.J., Rumbaugh, D.M., & Savage-Rumbaugh, E.S. (1998). Chimpanzee (*Pan troglodytes*) counting in a computerized testing paradigm. *The Psychological Record, 48,* 3-19.

Bernstein, I.S. (1961). Response variability and rigidity in the adult chimpanzee. *Journal of Gerontology, 16,* 381-386.

Biro, D., & Matsuzawa, T. (1999). Numerical ordering in a chimpanzee (*Pan troglodytes*): planning, executing, and monitoring. *Journal of Comparative Psychology, 113,* 178-185.

Birren, J.E., & Botwinick, J. (1951). The rate of addition as a function of difficulty and age. *Psychometrica, 16,* 219-232.

Bloomstrand, M.A. (1985). *An analysis of learning and memory in two aged chimpanzees.* Unpublished master's thesis, Georgia Institute of Technology, Atlanta, GA.

Borkhuis, M.L., Davis, R.T., & Medin, D.L. (1971). Confusion errors in monkey short term memory. *Journal of Comparative and Physiological Psychology, 77,* 206-211.

Bowden, D.M., & Jones, M.L. (1979). Aging research in nonhuman primates. In D.M. Bowden (Ed.), *Aging in Nonhuman Primates* (pp. 1-13). New York, NY: Van Nostrand Reinhold Company.

Boysen, S.T. (1993). Counting in chimpanzees: Nonhuman principles and emergent properties of number. In S.T. Boysen & Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 39-59). Hillsdale, NJ: Lawrence Erlbaum Associates.

Boysen, S.T. (1997). Representation of quantities by apes. *Advances in the study of behavior, 26,* 435-462.

Boysen, S.T., & Berntson, G.G. (1989). Numerical competence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology, 103,* 23-31.

Boysen, S.T., Berntson, G.G., Hannan, M.B., & Cacioppo, J.T. (1996). Quantity-based inference and symbolic representations in chimpanzee (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 22,* 76-86.

Boysen, S.T., Berntson, G.G., Shreyer, T.A., & Quigley, K.S. (1993). Processing of ordinality and transitivity by chimpanzees (*Pan troglodytes). Journal of Comparative Psychology, 107,* 208-215.

Boysen, S.T., & Hallberg, K.I. (2000). Primate numerical competence: Contributions towards understanding nonhuman cognition. *Cognitive Science, 24(3),* 423-443.

Call, J. (2000). Estimating and operating on discrete quantities in orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology, 114(2),* 136-147.

Cohen, D., Eisdorfer, C., & Bowden, D.M. (1979). Cognition. In D.M. Bowden (Ed.), *Aging in Nonhuman Primates* (pp. 48-55). New York, NY: Van Nostrand Reinhold Company.

Cooper, R.G., Jr. (1984). Early number development: Discovering number space and addition and subtraction. In C. Sophian (Ed.), *Origin of Cognitive Skills* (pp. 157-192). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis, H. (1993). Numerical competence in animals: A conservative view. In S.T. Boysen & Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 109-125). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis, R.T. (1978). Old monkey behavior. *Experimental Gerontology, 13*, 237-250.

Davis, R.T. (1985). The effects of aging on the behavior of rhesus monkeys. In R.T. Davis & C.W. Leathers (Eds.), *Behavior and pathology of aging in rhesus monkeys*. New York, NY: Alan R. Liss.

Davis, R.T., Bennett, C.L., & Weisenburger, W.P. (1982). Repeated measurements of forgetting by rhesus monkeys (*Macaca mulatta*). *Perceptual and Motor Skills, 55,* 703-709.

Davis, H., & Pérusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral and Brain Sciences, 11*, 561-615.

Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: A case study of severe acalculia with preserved approximation. *Neuropsycholgia, 29(11),* 1045-1074.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representation of numbers in the animal and human brain. *Trends in Neurosciences, 21(8),* 355-361.

Dooley, G.G., & Gill, T. (1977). Mathematical capabilities of Lana Chimpanzee. In G. H. Bourne (Ed.), *Progress in Ape Research* (pp. 133-142). New York, NY: Academic Press.

Ferster, C. (1964). Arithmetic behavior in chimpanzee. *Scientific American, 210*, 98 -105.

Flicker, C., Bartus, R.T., Crook, T., & Ferris, S.H. (1984). Effects of aging and dementia upon recent visuospatial memory. *Neurobiology of Aging, 5,* 75-83.

Gallagher, M, & Rapp, P.R. (1997). The use of animal models to study the effects of aging on cognition. *Annual Review of Psychology, 48,* 339-370.

Gallistel, C.R. (1989). Animal cognition: The representation of space, time, and number. *Annual Review of Psychology, 40,* 155-189.

Geary, D.C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114,* 345-362.

Geary, D.C., Bow-Thomas, C.C., Liu, F., & Siegler, R.S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development, 67,* 2022-2044.

Geary, D.C., Frensch, P.A., & Wiley, J.G. (1993). Simple and complex mental subtraction: strategy choice and speed-of-processing differences in younger and older adults. *Psychology and Aging, 8,* 242-256.

Geary, D.C., Hamson, C.O., Chen, G.-P., Liu, F., Hoard, M.K., & Salthouse, T.A. (1997). Computational and reasoning abilities in arithmetic: Cross-generational change in China and the United States. *Psychonomic Bulletin & Review, 4,* 425-430.

Geary, D.C., & Lin, J. (1998). Numerical cognition: Age-related differences in the speed of executing biologically primary and biologically secondary processes. *Experimental Aging Research, 24(2),* 101-137.

Geary, D.C., Salthouse, T.A., Chen, G.-P., & Fan, L. (1996). Are East Asian versus American differences in arithmetical ability a recent phenomenon? *Developmental Psychology, 32,* 254-262.

Geary, D.C., & Wiley, J.G. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology and Aging, 6,* 474-483.

Hauser, M.D. (1997). Tinkering with minds from the past. In G.R. Brock & G. Cardew (Eds.), *Characterizing human psychological adaptations* (pp. 95-131). Chichester, NY: John Wiley and Sons.

Hayes, W.L. (1994). *Statistics* (5th ed.). New York, NY: Harcourt Brace College Publishers.

Hayes, K.J., & Nissen, C.H. (1971). The higher mental functions of a home-raised chimpanzee. In A.M. Schrier & F. Stollnitz (Eds.), *Behavior of Nonhuman Primates: Modern Research Trends* (4th ed., pp. 59-115). New York, NY: Academic Press.

Herndon, J.G., Moss, M.B., Rosene, D.L., & Killiany, R.J. (1997). Patterns of cognitive decline in aged rhesus monkeys. *Behavioural Brain Research, 87,* 25-34.

Johnson, C.K., & Davis, R.T. (1973). Seven-year retention of oddity learning set in monkeys. *Perceptual and Motor Learning, 37,* 920-922.

Kaufman, E.L., Lord, M.W., Reese, T.W., & Volkmann, J. (1984). The discrimination of visual number. American *Journal of Psychology, 62,* 498-525.

Kotary, L., & Hoyer, W.J. (1995). Age and the ability to inhibit distracter information in visual selective attention. *Experimental Aging Research, 21,* 159-171.

Lai, Z.C., Moss, M.B., Killiany, R.J., Rosene, D.L., & Herndon, J.G. (1995). Executive system dysfunction in the aged monkey: Spatial and object reversal learning. *Neurobiology of Aging, 16(6)*, 947-954.

LeFevre, J.-A., Sadesky, G.S., & Bisanz, J. (1996). Selection procedures in mental additions: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 216-230.

Mandler, G., & Shebo. N.J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General,* 111, 1-22.

Marriott, J.G., & Abelson, J.S. (1980). Age differences in short-term memory of test-sophisticated rhesus monkeys. *Age, 3*, 7-9.

Medin, D.L. (1969). Form perception and pattern reproduction in monkeys. *Journal of Comparative and Physiological Psychology, 68,* 412-419.

Menzel, E.W., Jr. (1960). Selection of food by size in the chimpanzee, and comparison with human judgment. *Science, 131,* 1527-1528.

Miller, D.J. (1993). Do animals subitize?. In S.T. Boysen & E.J. Capaldi (Eds.), *The Development of Numerical Competence: Animal and Human Models* (pp. 149-169). Hillsdale, NJ: Lawrence Erlbaum Associates.

Murofushi, K. (1997). Numerical matching behavior by a chimpanzee (*Pan troglodytes*): Subitizing and analogue magnitude estimation. *Japanese Psychological Research, 39 (3),* 140-153.

Nebes, R.D., Brady, C.B., & Reynolds, C.F., III. (1992). Cognitive slowing in Alzheimer's disease and geriatric depression. *Journal of Gerontology: Psychological Sciences, 47,* 331-336.

Olthof, A., Iden, C.M., & Roberts, W.A. (1997). Judgments of ordinality and summation of number symbols by squirrel monkeys (*Saimiri sciureus*). *Journal of Experimental Psychology, 23(3),* 325-339.

Pérusse, R. & Rumbaugh, D. (1990). Summation in chimpanzees: Effects of amount, number of wells and finer ratios. *International Journal of Primatology, 11,* 425-437.

Rapp, P.R. (1990). Visual discrimination and reversal learning in the aged monkey (Macaca mulatta). *Behavioral Neuroscience, 104*, 876-884.

Rapp, P.R. (1995). Cognitive neuroscience perspectives on aging in nonhuman primates. In T. Nakajima & T. Ono (Eds.), *Emotion, Memory and Behavior: Studies on Human and Nonhuman Primates* (pp. 197-211). Tokyo, Japan: Japan Scientific Societies Press and Boca Raton, FL: CRC Press.

Riopelle, A.J., & Rogers, C.M. (1965). Age changes in Chimpanzees. In A.M. Schrier, H.F. Harlow, and F. Stollnitz (Eds.), *Behavior of Nonhuman Primates: Modern Research Trends* (pp. 449-462). New York, NY: Academic Press.

Rumbaugh, D.M., Savage-Rumbaugh, E.S., & Hegel, M. (1987). Summation in the chimpanzee. *Journal of Experimental Psychology: Animal Behavior Processes, 13,* 107 -115.

Rumbaugh, D.M., Savage-Rumbaugh, E.S., & Pate, J.L. (1988). Addendum to "Summation in the chimpanzee (*Pan troglodytes)." Journal of Experimental Psychology: Animal Behavior Processes, 14(1),* 118-120.

Rumbaugh, D., & Washburn, D.A. (1983). Counting by chimpanzees and ordinality judgments by macaques in video-formatted tasks. In S.T. Boysen & E.J. Capaldi (Eds.), *The Development of Numerical Competence: Animal and Human Models* (pp. 87-106). Hillsdale, NJ: Lawrence Erlbaum Associates.

Salthouse, T.A., & Kersten, A.W. (1993). Decomposing adult age differences in symbol arithmetic. *Memory & Cognition, 21,* 699 -710.

Schaller, G.B. (1965). The behavior of the mountain gorilla. In I. DeVore (Ed.) *Primate behavior: Field studies of monkeys and apes* (pp. 332-367). New York, NY: Holt, Rinehart and Winston.

Shumaker, R.W., Palkovich, A. M., Beck, B. B., Guagnano, G. A., & Morowitz, H. (2001). Spontaneous use of magnitude discrimination and ordination by the orangutan (*Pongo pygmaeus*). *Journal of Comparative Psychology, 115(4),* 385-391.

Silberberg, A., & Fujita, K. (1996). Pointing at smaller food amounts in an analogue of Boysen and Berntson's (1995) procedure. *Journal of Experimental Analysis of Behavior, 66,* 143-147.

Silwinski, M. (1997). Aging and counting speed: Evidence for process-specific slowing. *Psychology and Aging, 12,* 38-49.

Silwinski, M., Buschke, H., Kuslansky, G., Senior, G., & Scarisbrick, D. (1994). Proportional slowing and addition speed in old and young adults. *Psychology and Aging, 9,* 72-80.

Stevens, S.S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology* (pp. 1-49). New York, NY: John Wiley and Sons.

Tarou, L.R., Bloomsmith, M.A., Hoff, M.P., Erwin, J.M., & Maple, T.M. (2002). The behavior of aged great apes. In P.R. Hof & C.V. Mobbs (Series Eds.) & J.M. Erwin & P.R. Hof (Vol. Eds.), *Aging in Nonhuman Primates. Interdisciplinary topics in Gerontology* (Vol. 31, pp. 209-231). New York, NY: Karger Publishers.

Thomas, R.K., Fowlkes, D., & Vickery, J.D. (1980). Conceptual numerousness judgments by squirrel monkeys. *American Journal of Psychology, 93(2),* 247-257.

Thomas, R.K., & Lorden, R.B. (1993). Numerical competence in animals: A conservative view. In S.T. Boysen & E.J. Capaldi (Eds.), *The Development of Numerical Competence: Animal and Human Models* (pp. 127-147). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thomas, R.K., Phillips, J.A., & Young, C.D. (1999). Comparative cognition: Human numerousness judgments. *American Journal of Psychology, 112, (2),* 215-233.

Tomonaga, M., & Matsuzawa, T. (2002). Enumeration of briefly presented items by the chimpanzee (*Pan troglodytes*) and humans (*Homo sapiens*). *Animal Learning & Behavior, 30(2),* 143-157.

Trick, L.M., Enns, J.T., & Brodeur, D.A. (1996). Life span changes in visual enumeration: The number discrimination task. *Developmental Psychology, 32,* 925-932.

von Glaserfeld, E. (1993). Reflections on number and counting. In S.T. Boysen & E.J. Capaldi (Eds.), *The Development of Numerical Competence: Animal and Human Models* (pp. 225-243). Hillsdale, NJ: Lawrence Erlbaum Associates.

Walker, L.C., Kitt, C.A., Struble, R.G., Wagster, M.V., Price, D.L., & Cork, L.C. (1988). The neural basis of memory decline in aged monkeys. *Neurobiology of Aging, 9,* 657-666.

Washburn, S.L. (1981). Longevity in primates. In J.L. McGaugh & S.B. Kiesler (Eds.), *Aging: Biology and Behavior* (pp. 11-29). New York, NY: Academic Press.

Washburn, S.L., & Rumbaugh D.M. (1991). Ordinal judgments of numerical symbols by macaques (*Macaca mulatta*). *Psychological Science, 2,* 190-193.

Watson, D.G., Maylor, E.A., & Manson, N.J. (2002). Aging and enumeration: A selective deficit for the subitization of targets among distractors. *Psychology and Aging, 17(3),* 496-504.

Woodruff, G., & Premack, D. (1981). Primitive mathematical concepts in the chimpanzee: Proportionality and numerosity. *Nature, 293,* 568-570.

Woodruff-Pak, D.S. (1990). Mammalian models of learning, memory, and aging. In J.E. Birren and K.W. Schaie (Eds.), *Handbook of the Psychology of Aging* (3rd ed., pp. 201-221). San Diego, CA: Academic Press.

Wynn, K. (1998). An evolved capacity for number. In D.D. Cummins & C. Allen (Eds.), *The Evolution of Mind* (pp. 107-126). New York, NY: Oxford University Press.